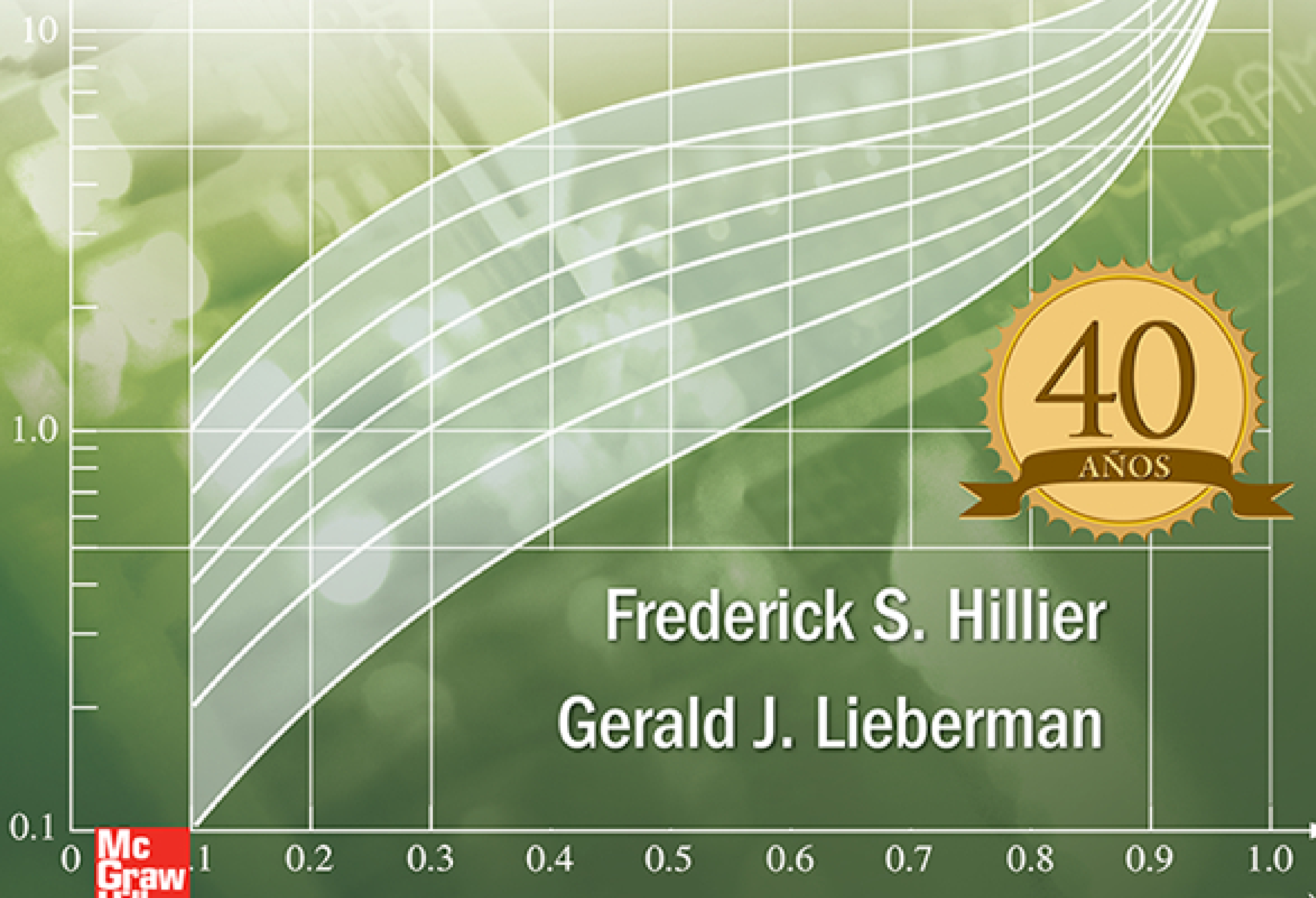
Introducción a la

INVESTIGACIÓN DE OPERACIONES

Novena edición



Frederick S. Hillier
Gerald J. Lieberman

Mc Graw Hill

INTRODUCCIÓN A LA INVESTIGACIÓN DE OPERACIONES

INTRODUCCIÓN A LA INVESTIGACIÓN DE OPERACIONES

Novena edición

Frederick S. Hillier
Stanford University

Gerald J. Lieberman
Late of Stanford University

Revisión técnica

Guillermo Martínez del Campo V.
Universidad Iberoamericana, Ciudad de México

Ernesto A. Pacheco
Instituto Tecnológico y de Estudios Superiores de Monterrey
Campus Ciudad de México

McGraw Hill

MÉXICO • BOGOTÁ • BUENOS AIRES • CARACAS • GUATEMALA • MADRID • NUEVA YORK
SAN JUAN • SANTIAGO • SÃO PAULO • AUCKLAND • LONDRES • MILÁN • MONTREAL
NUEVA DELHI • SAN FRANCISCO • SINGAPUR • ST. LOUIS • SIDNEY • TORONTO

Director Higher Education: Miguel Ángel Toledo Castellanos
Editor sponsor: Pablo E. Roig V.
Coordinadora editorial: Marcela I. Rocha Martínez
Editor de desarrollo: Edmundo Carlos Zúñiga Gutiérrez
Supervisor de producción: Zeferino García García
Traductores: Jesús Elmer Murrieta Murrieta y Carlos Roberto Cordero Pedraza

INTRODUCCIÓN A LA INVESTIGACIÓN DE OPERACIONES
Novena edición

A Subsidiary of ***The McGraw-Hill*** *Companies, Inc.*
Prolongación Paseo de la Reforma 1015, Torre A,
Piso 17, Colonia Desarrollo Santa Fe,
Delegación Álvaro Obregón,
C.P. 01376, México, D.F.
Miembro de la Cámara Nacional de la Industria Editorial Mexicana, Reg. Núm. 736

ISBN: 978-607-15-0308-4
(ISBN edición anterior: 970-10-5621-3)

Traducido de la novena edición de *Introduction to operations research,* by Frederick S. Hiller and Gerald J. Lieberman

0-07-337629-9

1234567890 109876543210

Impreso en México *Printed in Mexico*

ACERCA DE LOS AUTORES

Frederick S. Hillier nació y creció en Aberdeen, Washington, donde ganó premios estatales en concursos escolares de nivel medio en elaboración de ensayos, matemáticas, debate y música. Como estudiante en la Universidad de Stanford fue el primero en su clase de Ingeniería entre 300 estudiantes. También ganó el premio McKinsey por sus artículos técnicos, ganó el premio de debate para estudiantes de segundo año, tocó en el quinteto de aire de Stanford y ganó el premio Hamilton por combinar la excelencia en Ingeniería con logros notables en Humanidades y Ciencias Sociales. Después de su graduación en Ingeniería Industrial, fue premiado con tres becas nacionales (National Science Foundation, Tau Beta Pi y Danforth) para realizar sus estudios de posgrado en Stanford con especialización en Investigación de Operaciones. Después de recibir su grado de Doctor, se unió al cuerpo de profesores de la Universidad de Stanford, donde alcanzó el grado de profesor asistente a los 28 años y de profesor de tiempo completo a los 32 años. También fue maestro visitante en la Universidad de Cornell, Universidad Carnegie-Mellon, Universidad Técnica de Dinamarca, Universidad de Canterbury (Nueva Zelanda) y la Universidad de Cambridge (Inglaterra). Después de 35 años en Stanford, tomó un retiro voluntario de sus responsabilidades en el magisterio en 1996 para enfocarse de tiempo completo en la autoría de libros, y ahora es profesor emérito en Investigación de Operaciones en Stanford.

La investigación del Dr. Hillier se ha extendido a una gran variedad de áreas, entre las cuales se incluyen Programación Entera, Teoría de Colas y su aplicación, Control Estadístico de la Calidad y Aplicación de la Investigación de Operaciones en el Diseño de Sistemas Productivos y de Presupuestos de Capital. Ha publicado de manera continua y sus documentos de seminario han sido seleccionados para su publicación en libros de lecturas selectas al menos diez veces. Fue ganador del primer premio del concurso de investigación en "Presupuestos de Capital para Proyectos Interrelacionados" patrocinado por The Institute of Management Science (TIMS) y la Oficina de Investigación Naval de Estados Unidos. Junto con el Dr. Lieberman fue reconocido con la mención honorífica del premio Lanchester de 1995 (mejor publicación en inglés de cualquier tipo en el campo de la Investigación de Operaciones), que le fue otorgado por el Institute of Operations Research and the Management Sciences (INFORMS) por la sexta edición de este libro. Asmismo, ganó el prestigiado premio 2004 INFORMS Expository Writing Award por la octava edición de este libro.

El Dr. Hillier ha desempeñado muchos puestos de liderazgo en las sociedades profesionales de su campo. Por ejemplo, ha servido como tesorero de la Operations Research Society of America (ORSA), vicepresidente de reuniones de TIMS, codirector general de la Reunión Internacional de 1989 de TIMS en Osaka, Japón, director del Comité de Publicaciones de TIMS, director del Comité de búsqueda de editor en Investigación de Operaciones de ORSA, director del Comité de planeación de recursos de ORSA, director del Comité de reuniones combinadas ORSA/TIMS y director del Comité de selección para el premio en teoría John von Neumann de INFORMS. En la actualidad trabaja como editor de la International Series in Operations Research and Management Science que publica Springer, conformada por una relevante serie de libros que él fundó en 1993.

Además de *Introduction to Operations Research* y de los dos volúmenes que lo acompañan, *Introduction to Mathematical Programming* (2a. ed., 1995) e *Introduction to Stochastic Models in Operations Research* (1990), sus libros son *The Evaluation of Risky Interrelated Investments* (North-Holland, 1969), *Queueing Tables and Graphs* (Elsevier North-Holland, 1981, coescrito con O. S. Yu, con D. M. Avis, L. D. Fossett, F. D. Lo y M. I. Reiman), e *Introduction to Management Science: A Modeling and Case Studies Approach with Spreadsheets* (3a. ed., McGraw-Hill/Irwin, 2008, coescrito con M. S. Hillier).

Desafortunadamente, el otro autor, **Gerald J. Lieberman**, falleció en 1999. Fue profesor emérito en Investigación de Operaciones y Estadística de la Universidad de Stanford, donde fue director fundador del Departamento de Investigación de Operaciones. Fue ingeniero (pues recibió un grado en Ingeniería Mecánica de Cooper Union) y estadístico en Investigación de Operaciones

(con una maestría de la Universidad de Columbia en Estadística Matemática y un doctorado de la Universidad de Stanford en estadística).

El Dr. Lieberman fue uno de los líderes más eminentes de Stanford en las décadas recientes. Después de dirigir el Departamento de Investigación de Operaciones, fue decano asociado de la Escuela de Humanidades y Ciencias, vicerrector y decano de investigación, vicerrector y decano de estudios de posgrado, director del Senado de la facultad, miembro del Consejo de la Universidad y del Comité de celebración del centenario. También trabajó en Stanford como rector o rector activo bajo tres diferentes presidentes de esta institución.

Durante estos años de liderazgo en la universidad, también permaneció activo profesionalmente. Su investigación se destacó en las áreas estocásticas de Investigación de Operaciones, con frecuencia en la interfase de la Probabilidad Aplicada y la Estadística. Publicó de manera extensa en las áreas de Confiabilidad y Control de Calidad y en el Modelado de Sistemas Complejos, en el cual incluyó su Diseño Óptimo cuando los recursos son limitados.

El Dr. Lieberman, que fue reconocido como uno de los líderes más importantes en el campo de la Investigación de Operaciones, desempeñó numerosos papeles de liderazgo, como presidente electo del Institute of Management Sciences. Sus honores profesionales incluyeron ser elegido a la National Academy of Engineering, donde recibió la medalla Shewhart de la American Society for Quality Control, recibió el premio Cuthbertson por su servicio excepcional a la Universidad de Stanford y trabajó como adjunto en el centro para estudios avanzados en Ciencias del Comportamiento. Además, el Institute of Operations Research and the Management Sciences (INFORMS) lo premió junto al Dr. Hillier con la mención honorífica del premio Lanchester de 1995 por la sexta edición de este libro. En 1996, INFORMS también lo premió con la prestigiosa medalla Kimball por sus contribuciones excepcionales al campo de la Investigación de Operaciones y Ciencias de la Administración.

Además de *Introduction to Operations Research* y los dos volúmenes que lo acompañan, *Introduction to Mathematical Programming* (2a. ed., 1995) e *Introduction to Stochastic Models in Operations Research* (1990), sus libros son *Handbook of Industrial Statistics* (Prentice-Hall, 1955, coescrito con A. H. Bowker), *Tables of the Non-Central t-Distribution* (Stanford University Press, 1957, coescrito con G. J. Resnikoff), *Tables of the Hypergeometric Probability Distribution* (Stanford University Press, 1961, coescrito con D. Owen), *Engineering Statistics*, segunda edición (Prentice-Hall, 1972, coescrito con A. H. Bowker), e *Introduction to Management Science: A Modeling and Case Studies Approach with Spreadsheets* (McGraw-Hill/Irwin, 2000, coescrito con F. S. Hillier y M. S. Hillier).

ACERCA DE LOS AUTORES DE CASOS

Karl Schmedders es profesor asociado en el Departamento de Economía Administrativa y Ciencias de la Decisión en la Escuela de posgrado Kellogg de Administración (Northwestern University), donde enseña Métodos Cuantitativos para la Toma de Decisiones Administrativas. Sus intereses en investigación incluyen aplicaciones de la Investigación de Operaciones en la Teoría Económica, Teoría del Equilibrio General con Mercados Imperfectos, Precio de Activos y Economía Computacional. El Dr. Schmedders recibió su doctorado en Investigación de Operaciones de la Universidad de Stanford, donde impartió cursos de esta disciplina a nivel licenciatura y posgrado. Entre las clases que impartió se puede mencionar el curso sobre casos de estudio en investigación de operaciones, y después fue invitado a dictar una conferencia auspiciada por el Institute of Operations Research and the Management Sciences (INFORMS) acerca de su experiencia exitosa con este curso. Recibió diversos premios por su enseñanza en Stanford, que incluyen el prestigiado reconocimiento de la Universidad Walter J. Gores Teaching Award. También fue nombrado profesor del año L. G. Lavengood en la Escuela de Administración Kellog. Después de impartir cursos como profesor visitante en WHU Koblenz (una escuela de negocios líder en Alemania), ganó varios premios a la enseñanza.

Molly Stephens es asociada en la oficina de Los Angeles de Quinn, Emanuel, Urquhart, Oliver & Hedges, LLP. Se graduó en la Universidad de Stanford con una licenciatura en Ingeniería Industrial y una maestría en Investigación de Operaciones. La señora Stephens fue profesora de la Escuela de Ingeniería en Stanford y sirvió como asistente de enseñanza en el curso de casos de estudio en esta disciplina. Como asistente de enseñanza analizó problemas de Investigación de Operaciones del mundo real y la transformación de estos problemas en estudios de caso para el salón de clases. Su investigación fue premiada con una beca de investigación para estudiantes de Stanford, donde continuó con su trabajo y fue invitada a participar en una conferencia de INFORMS para presentar sus conclusiones sobre estudios de casos exitosos en el salón de clases. Después de su graduación, la señora Stephens trabajó en Andersen Consulting como integradora de sistemas, donde experimentó casos reales desde adentro, antes de continuar sus estudios de posgrado en la escuela de leyes de la Universidad de Texas en Austin donde obtuvo el grado con honores.

DEDICATORIA

A la memoria de nuestros padres

y

a la memoria de mi querido mentor,
Gerald J. Lieberman, quien fue un verdadero gigante
en nuestro campo de estudio

CONTENIDO

CAPÍTULO 18

Teoría de inventarios 772

CAPÍTULO 19

Procesos de decisión markovianos 844

CAPÍTULO 20

Simulación 871

COMPLEMENTOS DISPONIBLES EN EL SITIO EN INTERNET DE ESTE LIBRO www.mhhe.com/hillier

CASOS ADICIONALES

Caso 3.2 Disminución de costos de una cafetería
Caso 3.3 Asignación de personal en un centro de llamadas
Caso 3.4 Promoción de un cereal para el desayuno
Caso 4.2 Nuevas fronteras
Caso 4.3 Asignación de estudiantes a escuelas
Caso 6.2 Administración de granjas
Caso 6.3 Asignación de estudiantes a escuelas (revisado)
Caso 6.4 Redacción de una síntesis ejecutiva
Caso 8.2 Continuación del caso de estudio de Texago
Caso 8.3 Elección de proyectos
Caso 9.2 Ayuda a los aliados
Caso 9.3 Pasos hacia el éxito
Caso 11.2 Asignación de arte
Caso 11.3 Juegos de cocina en almacén
Caso 11.4 Asignación de estudiantes a escuelas (nueva revisión)
Caso 12.2 Inversiones internacionales
Caso 12.3 Promoción de un cereal para el desayuno, revisado
Caso 15.2 Apoyo inteligente al conductor
Caso 15.3 ¿Quién quiere ser millonario?
Caso 15.4 La universidad de los juguetes y los hombres de acción del profesor de ingeniería
Caso 17.2 Dilema de las colas
Caso 18.2 Aprovechamiento de las enseñanzas del voceador
Caso 18.3 Descarte el inventario excedente
Caso 20.3 Planeación de aplanadores
Caso 20.4 Precios bajo presión

COMPLEMENTO 1 DEL CAPÍTULO 3

El lenguaje de modelación de LINGO

COMPLEMENTO 2 DEL CAPÍTULO 3

Más acerca de LINGO

COMPLEMENTO DEL CAPÍTULO 7

Programación lineal de objetivos y sus procedimientos de solución

Problemas
Caso 7S.1 Una cura para Cuba
Caso 7S.2 Seguridad en aeropuertos

COMPLEMENTO DEL CAPÍTULO 8

Caso de estudio con muchos problemas de transporte

PREFACIO

Cuando Jerry Lieberman y yo comenzamos a trabajar en la primera edición de este libro hace 45 años, nuestra meta era desarrollar un libro de texto reformador que ayudara a establecer la dirección futura de la enseñanza de lo que entonces era el campo emergente de la Investigación de Operaciones. Después de la publicación, no fue claro en qué medida habíamos alcanzado ese objetivo, pero lo que sí quedó claro fue que la demanda por el libro era mucho más grande de lo que cualquiera de nosotros había anticipado. Nadie podría haber imaginado que esta extensa demanda a lo ancho del mundo continuaría a un nivel tan alto por un periodo tan largo.

La respuesta entusiasta a nuestras primeras ocho ediciones ha sido gratificante. Una satisfacción particular fue obtener de la sociedad profesional más importante en este campo, el Instituto Internacional de Investigación de Operaciones y Ciencias Administrativas (INFORMS), la mención honorífica para la sexta edición por el premio Lanchester 1995 (el premio otorgado para la publicación en inglés más destacada en el campo de la Investigación de Operaciones).

Después, una vez publicada esta octava edición, fue algo particularmente gratificante recibir el prestigiado premio 2004 INFORMS Expository Writing Award por este libro, que incluía la siguiente cita:

> Por más de 37 años, las ediciones sucesivas de este libro han generado que más de medio millón de estudiantes haya ingresado a este campo y han atraído a muchas personas a este campo de la actividad académica y la práctica profesional. Un gran número de líderes en esta área del conocimiento y muchos instructores aprendieron acerca de esta disciplina través de alguna edición de este libro. El uso intensivo de las ediciones internacionales para estudiantes y su traducción a 15 idiomas han contribuido a diseminarla alrededor del mundo. El libro mantiene su lugar en la vanguardia aún después de 37 años. A pesar de que acaba de salir la octava edición, la séptima acaparó 46% del mercado de libros de este tipo, y ocupó el segundo lugar en ventas internacionales con respecto a todas las publicaciones de McGraw-Hill en ingeniería.
>
> Para obtener este éxito, dos aspectos fueron muy importantes. En primer lugar, las ediciones han sido sorprendentes desde el punto de vista de los estudiantes debido a su excelente motivación, sus explicaciones claras e intuitivas, sus excelentes ejemplos acerca de la práctica profesional, a la excelente organización del material que presentan, a su software de soporte de gran utilidad y al uso moderado de las matemáticas. En segundo lugar, las ediciones han sido muy atractivas desde el punto de vista de los profesores debido a que han incluido de manera repetida material muy actual con una lucidez asombrosa y en un lenguaje muy claro. Por ejemplo, en la octava edición se añadió un fascinante capítulo acerca de metaheurística.

Cuando comenzamos a escribir el libro, hace 45 años, Jerry ya era un miembro prominente del campo, un autor de libros exitoso y director de un renombrado programa de Investigación de Operaciones en la Universidad de Stanford. Yo era un profesor asistente muy joven que apenas iniciaba mi carrera. Para mí fue una oportunidad maravillosa trabajar con y aprender del maestro. Siempre estaré en deuda con Jerry por haberme dado esta oportunidad.

Desafortunadamente, Jerry ya no está con nosotros. Durante la enfermedad progresiva que lo condujo a la muerte hace casi nueve años, decidí que tomaría la estafeta y me dedicaría por completo a las ediciones siguientes de este libro, que mantendría un estándar que honraría a cabalidad a Jerry. Por lo tanto, tomé un retiro adelantado de mis responsabilidades en Stanford con el fin de trabajar de tiempo completo en la redacción del libro para el futuro cercano. Este arreglo me ha permitido dedicarle más tiempo del usual a la preparación de esta nueva edición. También he podido supervisar de cerca las nuevas tendencias y desarrollos en el campo, para así lograr una edición completamente actualizada. Este monitoreo ha conducido a la selección de los nuevos temas que se describen a continuación.

CAMBIOS MÁS IMPORTANTES

- **Un gran hincapié en las aplicaciones reales.** Sin que el público en general lo sepa, el campo de la investigación de operaciones incrementa día a día su aportación al éxito de un gran número de compañías y organizaciones alrededor del mundo. Por lo tanto, un objetivo especial de esta edición es destacar la importancia de este libro y, por ende, hacer conscientes a los estudiantes de la gran relevancia del material que tienen en sus manos. Se ha tratado de alcanzar este objetivo de cuatro formas. La primera es *la adición de 29 viñetas de aplicación* separadas del material regular del texto que describen en unos cuantos párrafos la forma en que la aplicación de esta disciplina ha tenido un enorme efecto en una organización mediante el uso de técnicas como las que se estudian en esa parte específica del libro. La segunda es *la adición de 71 referencias seleccionadas de aplicaciones de IO ganadoras de premios* que se proporcionan al final de varios capítulos. La tercera es *la adición de una liga hacia los artículos que describen totalmente estas 100 aplicaciones*, a través de un arreglo especial con INFORMS. La última es *la incorporación de un gran número de problemas que requieren de la lectura de uno o más artículos*. Por lo tanto, el profesor puede motivar sus clases haciendo que sus alumnos se compenetren con las aplicaciones reales que demuestran de manera dramática la relevancia del material que se estudia en clase.

  Los autores están realmente emocionados acerca de su asociación con INFORMS, la sociedad profesional más importante en nuestro campo, para proporcionar un enlace a estos 100 artículos que describen aplicaciones dramáticas de la IO. El Instituto de Investigación de Operaciones y Ciencias Administrativas (INFORMS) es una sociedad profesional de estudiantes, académicos y practicantes en los campos cuantitativo y analítico. Para consultar información acerca de publicaciones periódicas, reuniones, bolsa de trabajo, becas, premios y material de enseñanza, ingrese a la página www.informs.org.
- **Aproximadamente 200 problemas nuevos o revisados.** Los problemas nuevos incluyen situaciones acerca de las aplicaciones reales que se mencionaron. Se han añadido nuevos problemas, entre ellos un considerable número de ellos que dan base a los temas nuevos y revisados que se mencionan más adelante. Se han agregado dos nuevos casos en el capítulo acerca del análisis de decisiones que tienen un menor grado de complejidad que los dos que ya estaban en él. Además, un gran número de problemas de la octava edición se han revisado. Por lo tanto, el profesor que no desee asignar problemas que ya se han asignado en clases anteriores, cuenta con una gran cantidad de éstos de dónde escoger.
- **Actualización del software que acompaña este libro.** La siguiente sección describe una gran cantidad de opciones de software que se proporciona en esta nueva edición. La diferencia principal con respecto a la octava edición radica en que ahora se encuentran disponibles versiones actualizadas de varios paquetes de software. Por ejemplo, *Excel 2007* representa la revisión más importante de Excel y su interfase de usuario por muchos años, por lo que esta nueva versión de Excel y su solucionador se han integrado a este libro (a la vez que se hace hincapié en las diferencias para facilitar la tarea de aquellas personas que todavía utilizan las versiones anteriores). Otro ejemplo importante es que, por primera vez en 10 años, se encuentran disponibles nuevas versiones de *TreePlan* y *SensIt*, totalmente integradas en el capítulo dedicado a la toma de decisiones. En esta nueva edición se proporcionan las últimas versiones de los demás paquetes de software.
- **Nueva sección acerca de la administración de los ingresos.** Un parte clave de este libro ha sido la adición del estudio de desarrollos recientes que han comenzado a revolucionar la forma en que se practican ciertas áreas de la investigación de operaciones. Por ejemplo, la octava edición agregó un capítulo acerca de metaheurística, una nueva sección acerca de la incorporación de la programación de restricciones y otra acerca de los modelos de inventarios multigrado para la administración de la cadena de suministro. Esta edición agrega otro tema fundamental: *una sección completa acerca de la administración de los ingresos en el capítulo acerca de la teoría de inventarios*. Este agregado representa un avance muy oportuno debido al enorme efecto que este tema ha tenido en la industria aeronáutica y que ha empezado a tener en otros sectores productivos.
- **Reorganización del capítulo acerca de la teoría del método símplex.** Algunos profesores no tienen tiempo para cubrir el método símplex revisado, pero puede ser que deseen presentar su

forma matricial y lo que llamamos la "idea fundamental" que se encuentra en sus cimientos. Por lo tanto, en lugar de cubrir el método símplex revisado de la seccion 5.2 antes de estudiar la idea fundamental en la sección 5.3 —como el la octava edición—, ahora sólo presentamos la forma matricial del método símplex en la sección 5.2, la cual fluye directamente de la idea fundamental de la sección 5.3, después de la cual nos enfocamos en el método símplex revisado como un tema opcional en la sección 5.4.

- **Método simplificado para determinar las utilidades.** Entre las demás revisiones sutiles de este libro, quizás la más importante sea la presentación simplificada en la sección 15.6 sobre cómo determinar utilidades, objetivo que se logra mediante la descripción de un "método de lotería equivalente".
- **Reorganización para reducir el tamaño del libro.** Una tendencia poco afortunada de las ediciones anteriores de este libro fue que cada nueva versión era mucho mayor que la anterior. Este sesgo continuó hasta que la séptima edición llegó a alcanzar un tamaño mucho más grande que el deseable para un libro de texto introductorio. Por lo tanto, se ha trabajado intensamente para reducir de manera sustancial las dimensiones de la octava edición y alcanzar el objetivo de evitar que el tamaño de las siguientes ediciones creciera. Dicho objetivo se cumple en esta novena edición a través de varias formas, una de las cuales es no agregar demasiado material. Otra fue eliminar dos secciones acerca de aplicaciones reales de la octava edición pero que ya no eran necesarias debido a la adición de viñetas de aplicación. Una tercera consistió en mover tanto el apéndice 3.1 acerca del lenguaje de modelación LINGO como la sección acerca de la optimización con OptQuest a los complementos del libro en su sitio en internet. (Esta decisión respecto de OptQuest se tomó con facilidad debido a que una nueva versión cae en la obsolescencia con rapidez, y la nueva versión no salió antes de la liberación de esta edición, por lo que se añadirá después como complemento). Por último, se redujo un número considerable de secciones. Por el contrario, se ha tratado de adherir a lo que se supone se ha convertido en la ya familiar organización de la octava edición después de introducirle cambios significativos.
- **Actualización con el fin de reflejar lo más actual en este campo.** Se ha hecho un esfuerzo especial para mantener actualizado este libro. Esto implica la actualización cuidadosa de las referencias seleccionadas al final de cada capítulo, así como también de los pies de página que hacen referencia a los últimos avances acerca de los temas en estudio.

GRAN CANTIDAD DE OPCIONES DE SOFTWARE

Se ofrece una enorme cantidad de software en el sitio en internet de este libro, www.mhhe.com/hillier, que se describe a continuación.

- Hojas de cálculo de Excel: se despliegan formulaciones actuales en hojas de cálculo en archivos de Excel para presentar los ejemplos relevantes a lo largo del libro.
- Varios complementos de Excel, entre ellos el Premium Solver for Education (una mejora del Solucionador básico de Excel), TreePlan (para el análisis de decisiones), SensIt (para el análisis estadístico de sensibilidad), RiskSim (para simulación) y Solver table (para el análisis de sensibilidad).
- En gran número de plantillas de Excel para resolver modelos básicos.
- Las versiones del estudiantes de LINDO (un optimizador convencional) y LINGO (un lenguaje de modelación algebraica muy popular), junto con las formulaciones y soluciones de todos los ejemplos relevantes del libro.
- Las versiones del estudiante de MPL (un lenguaje de modelación algebraica muy eficiente) y su solucionador principal CPLEX (el optimizador que más se utiliza) junto con el Tutorial MPL y formulaciones y soluciones MPL/CPLEX de todos los ejemplos relevantes del libro.
- Las versiones del estudiante de varios solucionadores MPL adicionales, entre ellos CONOPT (para la programación convexa), LGO (para la optimización global), LINDO (para la programación matemática), CoinMP (para la programación lineal y entera) y BendX (para algunos modelos estocásticos).
- Simulador de colas (para la simulación de sistemas de colas).
- OR Tutor para ilustrar varios algoritmos en acción.

- Tutorial de Investigación de Operaciones Interactivo (IOR) para aprender y ejecutar algoritmos interactivamente de manera eficiente, implantado en Java 2 con la finalidad de trabajar de forma independiente de la plataforma.

Un gran número de estudiantes consideran a OR Tutor y al Tutorial IOR como una herramienta muy valiosa para aprender los algoritmos de investigación de operaciones. De acuerdo con ciertos estudios, cuando pasan de manera automática a la siguiente etapa de resolución de modelos de IO, los profesores se inclinan de manera equitativa porque sus alumnos utilicen alguna de las opciones siguientes: 1) Hojas de cálculos de Excel, entre ellos el Excel Solver y otros complementos, 2) software tradicional (LINDO y LINGO) y 3) software para IO actual (MPL y CPLEX). Por lo tanto, en esta edición, se mantiene la filosofía de las últimas dos ediciones de proporcionar una introducción suficiente en el libro que permita el uso básico de cualquiera de las tres opciones sin distraer a los que usan otra opción, a la vez que también se proporciona una gran cantidad de material de soporte para cada opción en el sitio en internet del libro.

Se ha decidido no incluir el paquete de software Crystal Ball que contenía la octava edición. Por fortuna, en la actualidad muchas universidades cuentan con una licencia en el sitio de Crystal Ball y el paquete se puede descargar por un periodo de prueba de 30 días, por lo que es factible aún encontrarse con estudiantes que utilizan este software, al menos por un tiempo limitado. En consecuencia, esta edición continúa usando Cystal Ball en la sección 20.6 y ciertos complementos para ilustrar la sorprendente funcionalidad para analizar modelos de simulación actualmente disponibles.

Recursos adicionales en línea

- En la sección Worked Examples del sitio en internet de este libro se encuentran algunos ejemplos de casi todos los capítulos del libro con la finalidad de ayudar a los alumnos que lo necesiten sin romper con la continuidad del texto ni agregar páginas innecesarias a los demás. (En el libro se utilizan letras en negritas para resaltar siempre que un ejemplo adicional acerca del tema actual se encuentre disponible).
- Un glosario de cada capítulo del libro.
- Se incluyen archivos de datos de varios casos para permitir que los alumnos se enfoquen en el análisis de grandes conjuntos de datos en lugar de perder tiempo ingresándolos.
- Una abundante cantidad de texto complementario (ocho capítulos completos).
- Se les proporciona a los profesores un banco de datos que contiene preguntas con un moderado grado de dificultad para que los alumnos muestren su trabajo. Los autores han usado la mayoría de las preguntas que contiene dicho banco de datos con mucho éxito como preguntas de prueba.
- También se encuentran disponibles para los profesores el manual de respuestas y los archivos con imágenes.

Opción de libro de texto en formato electrónico

Este texto se ofrece a través de CourseSmart tanto para los profesores como para los estudiantes. CourseSmart es un recurso en línea donde los estudiantes pueden comprar el acceso a éste y otros libros de texto de McGraw-Hill en formato digital. A través de su navegador, los estudiantes pueden acceder en línea al texto completo a casi la mitad del costo de un libro convencional. La compra del libro de texto en formato electrónico también les permite aprovechar las herramientas en internet de CourseSmart para el aprendizaje, el cual incluye la herramienta para búsqueda de texto, notas y resalto y herramientas de correo electrónico para compartir notas entre compañeros de clase. Para aprender más acerca de las opciones de CourseSmart, contacte a su representante de ventas o visite el sitio www.CourseSmart.com.

USO DEL LIBRO

El objetivo global de todos los esfuerzos de revisión ha sido construir sobre las fortalezas de las ediciones anteriores para satisfacer en mayor medida las necesidades de los estudiantes de hoy. Tanto

los nuevos temas como las extensas actualizaciones hacen que el libro sea aún más adaptable para su uso en un curso moderno que refleje la práctica contemporánea en el campo. El uso del software es inmanente a la práctica de la investigación de operaciones, por lo que la abundancia de opciones de software que acompañan al libro proporciona una gran flexibilidad para que el profesor elija los tipos de software que deben usar los estudiantes. Todos los recursos educativos que acompañan al libro mejoran la experiencia de aprendizaje. Por lo tanto, el libro y su sitio en internet deben ajustarse a un curso donde el instructor quiere que los estudiantes tengan un solo libro de texto con contenido suficiente el cual complemente y dé apoyo a lo que sucede en el salón de clases.

El equipo editorial de McGraw-Hill y el autor piensan que el efecto neto de la revisión ha sido hacer esta edición aún más un “libro del estudiante”: claro, interesante y bien organizado con muchos ejemplos útiles e ilustraciones, buena motivación y perspectiva, con el material importante fácil de encontrar, tareas disfrutables, sin demasiada notación, terminología ni matemáticas densas. Creemos y confiamos en que los múltiples profesores que han utilizado las ediciones previas estarán de acuerdo en que ésta es la mejor edición hasta ahora.

Los prerrequisitos para un curso que utilice este libro pueden ser relativamente modestos. Como en las ediciones previas, las matemáticas se han mantenido a un nivel relativamente elemental. La mayoría de los capítulos del 1 al 14 (introducción, programación lineal y programación matemática) no requieren de matemáticas más allá del álgebra a nivel medio superior. El cálculo se utiliza sólo en los capítulos 12 (programación no lineal) y en un ejemplo del capítulo 10 (programación dinámica). La notación matricial se utiliza en el capítulo 5 (teoría del método símplex), capítulo 6 (teoría de dualidad y análisis de sensibilidad), en la sección 7.4 (algoritmo de punto interior) y en el capítulo 12 pero todo el respaldo necesario para ésta se presenta en el apéndice 4. Para los capítulos del 15 al 20 (modelos probabilísticos), se supone una introducción previa a la teoría de probabilidad y el cálculo se usa en unos cuantos sitios. En términos generales, la madurez matemática que logra un estudiante al tomar un curso elemental de cálculo es útil para los capítulos del 15 al 20 y para el material más avanzado que se incluye en los capítulos posteriores.

El contenido del libro es muy utilizado en la división superior del nivel de licenciatura (incluyendo alumnos de segundo año bien preparados) y en el primer año (a nivel maestría) de estudios de posgrado. Debido a la gran flexibilidad del libro hay muchas maneras de empaquetar el material en un curso. Los capítulos 1 y 2 dan una introducción a la materia de investigación de operaciones. Los capítulos del 3 al 14 (sobre programación lineal y programación matemática) pueden en esencia cubrirse independientemente de los capítulos 15 al 20 (sobre modelos probabilísticos), y viceversa. Aún más, cada uno de los capítulos entre el 3 y 14 son casi independientes, excepto porque todos ellos utilizan material básico que se presentó en el capítulo 3 y quizás en el 4. El capítulo 6 y la sección 7.2 también se basan en el capítulo 5. Las secciones 7.1 y 7.2 utilizan partes del capítulo 6. En la sección 9.6 se supone una concordancia con las formulaciones de problemas en las secciones 8.1 y 8.3, mientras que en la sección 9.7 resulta útil (pero no esencial) una exposición de las secciones 7.3 y 8.2. En los capítulos del 15 al 20 existe una flexibilidad de cobertura considerable, aunque se dispone de cierta integración del material.

Un curso introductorio elemental que cubra programación lineal, programación matemática y algunos modelos probabilísticos puede presentarse en un trimestre (40 horas) o un semestre al seleccionar en forma selectiva el material a lo largo del libro. Por ejemplo, una buena revisión del campo puede obtenerse de los capítulos 1, 2, 3, 4, 15, 17, 18 y 20, junto con partes de los capítulos 9 al 13. Un curso elemental más extenso se puede completar en dos trimestres (60 a 80 horas) si se excluyen sólo unos cuantos capítulos, por ejemplo 7, 14 y 19. Los capítulos 1 al 8 (y quizás una parte del capítulo 9) forman una base excelente para un curso (de un trimestre) en programación lineal. El material de los capítulos del 9 al 14 cubren temas para otro curso (de un trimestre) en modelos determinísticos. Por último, el material en los capítulos 15 al 20 cubren los modelos probabilísticos (estocásticos) de investigación de operaciones útiles para su presentación en un curso (de un trimestre). De hecho, estos tres últimos cursos (el material incluido en todo el texto) puede verse como una secuencia básica de un año en las técnicas de investigación de operaciones, lo que forma la esencia de un programa de maestría. Cada curso esquematizado se ha presentado a nivel licenciatura o posgrado en la Universidad de Stanford, y este texto se ha utilizado de la manera que se sugiere en ellos.

El sitio en internet proporcionará actualizaciones acerca del libro, incluyendo la fe de erratas. Para acceder a este sitio, visite www.mhhe.com/hillier

RECONOCIMIENTOS

Estoy en deuda con el excelente grupo de revisores que proporcionaron sabios consejos en el proceso de revisión. En este grupo se incluyen

Chun-Hung Chen, George Mason University
Mary Court, University of Oklahoma
Todd Easton, Kansas State University
Samuel H. Huang, University of Cincinnati
Ronald Giachetti, Florida International University
Mary E. Kurz, Clemson University
Wooseung Jang, University of Missouri-Columbia
Shafiu Jibrin, Northern Arizona University
Roger Johnson, South Dakota School of Mines & Technology
Emanuel Melachrinoudis, Northeastern University
Clark A. Mount-Campbell, The Ohio State University
Jose A. Ventura, Pennsylvania State University
John Wu, Kansas State University

También estoy muy agradecido con Garrett Ryzin por su consejo experto con respecto a la nueva sección acerca de la administración de los ingresos, a Charles McCallum Jr., por proporcionarme tres veces las listas de tipógrafos en la octava edición y a Bjarni Kristjansson por brindarme información actualizada acerca de los tamaños de los problemas que se estaban resolviendo con éxito por el software de optimización más actual. Además, quiero darle las gracias a los profesores y estudiantes que me enviaron mensajes de correo electrónico para darme su retroalimentación acerca de la octava edición.

Esta edición representa los esfuerzos de un equipo. Nuestros escritores de casos, Karl Schmedders y Molly Stephens (ambos con maestría y que trabajan en nuestro departamento), escribieron 24 casos para la séptima edición, todos los cuales forman parte de esta nueva edición. Una de las estudiantes de doctorado del departamento, Pelin Canbolat, hizo un excelente trabajo en la preparación del manual de respuestas. Ella superó las expectativas en su trabajo tecleando casi todas las respuestas que se hicieron en manuscrito en ediciones anteriores, así como aportando ideas en esta edición. Uno de nuestros estudiantes de doctorado anteriores, Michael O'Sullivan, desarrolló el OR Tutor para la séptima edición (y continúa todavía en el departamento), con base en parte del software que mi hijo Mark Hillier desarrolló para la quinta y sexta ediciones. Mark, (que nació el mismo año que la primera edición de este libro, obtuvo su doctorado en Stanford y, en este momento, es profesor asociado de métodos cuantitativos en la Universidad de Washington), proporcionó las hojas de cálculo y los archivos de Excel (incluyendo las plantillas de Excel) de esta edición, así como la tabla Solver y el simulador de colas. Asimismo, brindó consejos de gran utilidad acerca del material del texto y del software de esta edición y contribuyó de manera notable en los capítulos 21 y 28 en el sitio en internet de este libro. William Sun (presidente de la compañía de software Accelet Corporation), es otro graduado con doctorado en Stanford que junto con su equipo realizó un brillante trabajo al comenzar con gran parte del software que hizo Mark e implementarlo en Java 2 como parte del IOR Tutorial para la séptima edición. Estas personas hicieron un trabajo excelente que consistió en mejorar el IOR Tutorial para la octava edición y las subsecuentes. Linus Schrage, de la Universidad de Chicago y LINDO Systems (que tomó un curso introductorio de investigación de operaciones conmigo hace 45 años), proporcionó el software LINGO y LINDO para el sitio en internet de este libro. También supervisó el desarrollo de los archivos de LINGO/LINDO para los diferentes capítulos, y proporcionó material del tutorial para el sitio en internet de este libro. Otro antiguo amigo mío, Bjarni Kristjansson (director de la compañía Maximal Software), hizo lo mismo con los archivos MPL/CPLEX y el material del tutorial MPL, y también realizó los arreglos necesarios para proporcionar versiones del estudiante de MPL, CPLEX y otros solucionadores para el sitio en internet de este libro. Mi esposa, Ann Hillier, invirtió muchos días y noches sentada en una Macintosh capturando el texto y elaborando un gran número de figuras y tablas. Todos ellos fueron una parte vital del equipo.

Además de estar en deuda con Accelet Corporation, LINDO Systems y Maximal Software, estamos profundamente agradecidos con otras compañías que brindaron el software que acompaña

esta edición. Entre ellas se encuentran Frontline Systems (por proporcionar el Premium Solver for Education), ILOG (por brindar el solucionador CPLEX que se usó en la edición del estudiante de MPL), ARKI Corporation (por proporcionar el solucionador de programación convexa CONOPOT que se incluye en la edición del estudiante de MPL),y PCS Inc. (que proporcionó el solucionador de optimización global LGO que se utilizó en la versión del estudiante de MPL). También estamos muy agradecidos con el profesor Michael Middleton por brindar versiones actualizadas de TreePlan y SensIt, así como también de RiskSim. Por último, reconocemos la enorme ayuda de INFORMS, que proporcionó una liga hacia los artículos de la revista *Interfaces* que describen las aplicaciones de OR que se resumen en las viñetas de aplicación y en otras referencias seleccionadas de aplicaciones que se han hecho acreedoras a premios en el área de Investigación de Operaciones y que se brindan en este libro.

Para mí, fue un enorme placer trabajar con el equipo de producción editorial de McGraw-Hill, el que mostró mucho profesionalismo. En él se destacaron Debra Hash (editora sponsor) y Lora Kalb-Neyens (editora de desarrollo).

De la misma forma que un gran número de personas realizaron contribuciones importantes a esta edición, quisiera invitar a cada uno de quienes lean el texto a que comiencen a contribuir al nacimiento de la próxima edición usando mi dirección de correo electrónico que se encuentra abajo enviándome sus comentarios, sugerencias y los errores que detecten, a fin de mejorar este libro en el futuro. Al proporcionarles mi dirección de correo electrónico, me permito aclararles que continuaré aplicando la política de no enviar las respuestas a los problemas ni a los casos del libro a nadie que me contacte (incluyendo a sus alumnos).

Disfruten el libro.

Frederick S. Hillier
Stanford University

CAPÍTULO 1

Introducción

1.1 ORÍGENES DE LA INVESTIGACIÓN DE OPERACIONES

Desde el advenimiento de la Revolución industrial, el mundo ha sido testigo de un crecimiento importante del tamaño y la complejidad de las organizaciones. Los pequeños talleres artesanales de épocas anteriores se convirtieron en las corporaciones actuales de miles de millones de dólares. Una parte esencial de este cambio revolucionario fue el gran aumento de la división del trabajo y de la separación de las responsabilidades administrativas en estas organizaciones. Los resultados han sido espectaculares. Sin embargo, junto con los beneficios, el aumento del grado de especialización trajo consigo problemas nuevos que aún existen en numerosas organizaciones. Uno de éstos es la tendencia de algunos componentes de una organización a convertirse en imperios con autonomía relativa, con sus propias metas y sistemas de valores; de esta manera pierden de vista la forma en que sus actividades y objetivos se acoplan a los de toda la organización. Con frecuencia, lo que es mejor para un componente va en detrimento de otro, de forma que sus acciones pueden caminar hacia objetivos opuestos. Un problema relacionado es que, en la medida que aumentan la complejidad y la especialización, es más difícil asignar los recursos disponibles a las diferentes actividades de la manera más eficaz para la organización como un todo. Este tipo de problemas y la necesidad de encontrar la mejor forma de resolverlos crearon el ambiente propicio para el surgimiento de la **investigación de operaciones**, a la que también se hace referencia como **IO**.

Las raíces de la IO pueden encontrarse muchas décadas atrás,[1] cuando se hicieron los primeros intentos por emplear el método científico para administrar una empresa. Sin embargo, el inicio de la actividad llamada *investigación de operaciones* es atribuible a ciertos servicios militares que se prestaron al inicio de la Segunda Guerra Mundial. Debido a los esfuerzos bélicos, existía la urgente necesidad de asignar recursos escasos a las distintas maniobras militares y a las actividades que componían cada operación de la manera más eficaz. Por ello, las administraciones militares estadounidense y británica llamaron a un gran número de científicos para que aplicaran el método científico a éste y a otros problemas estratégicos y tácticos. En realidad, les solicitaron que hicieran *investigación sobre operaciones* (militares). Estos grupos de científicos fueron los primeros equipos de IO. Debido al desarrollo de métodos eficaces para utilizar la nueva herramienta que representaba el radar, los científicos contribuyeron al triunfo en la guerra aérea que libró Gran Bretaña. Sus investigaciones para mejorar el manejo de las operaciones antisubmarinas y de protección también tuvieron un papel importante en la victoria de la campaña del Atlántico Norte. Esfuerzos similares fueron de gran ayuda en la campaña del Pacífico.

Al terminar la guerra, el éxito de la IO en las actividades bélicas generó gran interés debido a las posibilidades de aplicarla en un ámbito distinto al militar. Una vez que la explosión industrial

[1] La referencia seleccionada 2 proporciona una entretenida historia de la investigación de operaciones que rastrea sus raíces hasta 1564 y describe una cantidad considerable de contribuciones científicas desde ese año hasta 1935 que influyeron en el desarrollo subsiguiente de la IO.

posterior a la guerra siguió su curso, los problemas provocados por el aumento de la complejidad y la especialización de las organizaciones pasaron de nuevo al primer plano. Entonces comenzó a ser evidente para un gran número de personas, entre ellas los consultores industriales que habían trabajado con o para los equipos de IO durante la guerra, que estos problemas eran en esencia los mismos que los que debían enfrentar los militares pero en un contexto diferente. Al inicio de la década de los años cincuenta, estos visionarios introdujeron el uso de la investigación de operaciones en una serie de organizaciones industriales, de negocios y del gobierno. Desde entonces, se ha desarrollado con rapidez.

Es posible identificar por lo menos otros dos factores que tuvieron gran importancia en el desarrollo de la IO durante este periodo. Uno es el progreso sustancial que se logró en el mejoramiento de las técnicas disponibles. Después de la guerra, muchos de los científicos que habían participado en equipos de IO o que tenían información sobre este trabajo, estaban motivados para realizar investigación relevante en el campo, de lo cual resultaron avances importantes; un ejemplo sobresaliente es el *método símplex* para resolver problemas de programación lineal, desarrollado en 1947 por George Dantzig. Muchas de las herramientas características de la IO, como programación lineal, programación dinámica, teoría de colas y teoría de inventarios habían sido desarrolladas casi por completo antes del término de la década de los años cincuenta.

Un segundo factor que dio gran impulso al desarrollo de este campo fue la *revolución de las computadoras.* El manejo eficaz de los complejos problemas inherentes a la IO casi siempre requiere un gran número de cálculos. Realizarlos de forma manual puede resultar casi imposible, por lo cual el desarrollo de la computadora electrónica digital, con su capacidad para hacer cálculos aritméticos, miles o tal vez millones de veces más rápido que los seres humanos, fue una gran ayuda para esta disciplina. Otro avance tuvo lugar en la década de los años ochenta, con el desarrollo de computadoras personales cada vez más rápidas y de buenos paquetes de software para resolver problemas de IO. De esta forma, las técnicas más complejas estuvieron al alcance de un gran número de personas, y este progreso se aceleró aún más en la década de 1990 y al inicio del siglo XXI. Hoy en día, millones de individuos tienen acceso a estos paquetes y en forma cotidiana se utiliza toda una gama de computadoras, desde las grandes hasta las portátiles, para resolver problemas de investigación de operaciones, algunos de ellos muy complejos.

1.2 NATURALEZA DE LA INVESTIGACIÓN DE OPERACIONES

Como su nombre lo indica, el objetivo de esta disciplina implica “investigar sobre las operaciones”. En consecuencia, esta disciplina se aplica a la problemática relacionada con la conducción y la coordinación de *actividades* en una organización. En esencia, la naturaleza de la organización es irrelevante, por lo cual la IO ha sido aplicada de manera extensa en áreas tan diversas como manufactura, transporte, construcción, telecomunicaciones, planeación financiera, cuidado de la salud, fuerzas armadas y servicios públicos, por nombrar sólo unas cuantas. Así, la gama de aplicaciones es inusualmente amplia.

La IO incluye el término *investigación* en el nombre porque utiliza un enfoque similar al que se aplica en las áreas científicas establecidas. El *método científico* se utiliza para explorar los diversos problemas que deben ser enfrentados, pero en ocasiones se usa el término *management science* o *ciencia de la administración* como sinónimo de investigación de operaciones. El proceso comienza por la observación cuidadosa y la formulación del problema, lo cual incluye la recolección de los datos pertinentes. El siguiente paso es la construcción de un modelo científico —generalmente matemático— con el cual se intenta abstraer la esencia del problema real. En esta etapa se propone la hipótesis de que el modelo será una representación tan precisa de las características esenciales de la situación, que permitirá que las conclusiones —soluciones— que se obtengan sean válidas también para el problema real. Después se llevan a cabo los experimentos adecuados para probar esta hipótesis, para modificarla si es necesario y para verificarla en determinado momento, paso que se conoce como *validación del modelo.* En cierto sentido, la IO involucra la investigación científica creativa de las propiedades fundamentales de las operaciones. Sin embargo, es más que esto. La IO se ocupa también de la administración práctica de la organización. Por lo tanto, para tener éxito, también debe proporcionar conclusiones claras que el tomador de decisiones pueda usar cuando sea necesario.

Otra característica de la investigación de operaciones es su amplio punto de vista. Como quedó implícito en la sección anterior, la IO adopta una visión organizacional. Desde esta perspectiva

intenta resolver los conflictos de intereses entre los componentes de la organización de forma que el resultado sea el mejor para ésta en su conjunto. Ello no significa que el estudio de cada problema deba considerar en forma explícita todos los aspectos de la organización, sino que los objetivos que se persiguen deben ser congruentes con los objetivos globales.

Una característica adicional de la investigación de operaciones es que intenta encontrar una *mejor* solución —llamada solución *óptima*— para el problema en cuestión. (Se dice *una* mejor solución y no *la* mejor solución porque es posible que existan muchas soluciones que puedan considerarse como las mejores.) En lugar de conformarse con mejorar el estado de las cosas, la meta es identificar el mejor curso de acción posible. Aun cuando debe interpretarse con todo cuidado en términos de las necesidades reales de la administración, esta "búsqueda del mejor camino" es un aspecto importante de la IO.

Estas características conducen de manera casi natural a otra. Es evidente que no puede esperarse que un solo individuo sea experto en los múltiples aspectos del trabajo de investigación de operaciones o de los problemas que se estudian, sino que se requiere un grupo de individuos con diversos antecedentes y aptitudes. Cuando se decide emprender un estudio de IO completo de un problema nuevo, es necesario emplear el *enfoque de equipo.* Este grupo de expertos debe incluir individuos con antecedentes sólidos en matemáticas, estadística y teoría de probabilidades, al igual que en economía, administración de empresas, ciencias de la computación, ingeniería, ciencias físicas, ciencias del comportamiento y, por supuesto, en las técnicas especiales de IO. El equipo también necesita experiencia y aptitudes necesarias para considerar de manera adecuada todas las ramificaciones del problema dentro de la organización.

1.3 EFECTO DE LA INVESTIGACIÓN DE OPERACIONES

La investigación de operaciones ha tenido un efecto impresionante en el mejoramiento de la eficiencia de numerosas organizaciones de todo el mundo. En el proceso, la IO ha contribuido de manera significativa al incremento de la productividad de la economía de varios países. Hoy existen más de 30 países miembros de la International Federation of Operational Research Societies (IFORS), cada uno de los cuales cuenta con una sociedad de investigación de operaciones. Tanto en Europa como en Asia existen federaciones de sociedades de IO que dan conferencias y publican revistas internacionales en esos continentes. Además, el Institute for Operations Research and the Management Sciences (INFORMS) es una sociedad de IO internacional. Entre sus múltiples revistas existe una, llamada *Interfaces*, que publica artículos que presentan estudios importantes de IO y el efecto que éstos tuvieron en sus organizaciones.

Para dar una mejor idea de la amplia aplicabilidad de la IO, en la tabla 1.1 se enumeran algunas aplicaciones reales. Observe la diversidad de organizaciones y aplicaciones incluidas en las primeras dos columnas. En la tercera columna se identifica la sección donde un "recuadro de aplicación" contiene varios párrafos en los que se describe la aplicación y también hace referencia a un artículo que proporciona detalles completos. (En esta sección se puede ver el primero de estos recuadros de aplicación.) La última columna indica que estas aplicaciones significaron ahorros anuales de muchos millones de dólares. Aún más, algunos beneficios adicionales no registrados en la tabla como un mejor servicio al cliente y mayor control administrativo fueron considerados más importantes, en ciertos casos, que los beneficios financieros. (El lector tendrá oportunidad de investigar estos beneficios menos tangibles en los problemas 1.3-1, 1.3-2 y 1.3-3.) En nuestro sitio web —www.mhhe.com/hillier— se incluye un vínculo con los artículos que describen estas aplicaciones a detalle.

Aunque la mayoría de los estudios rutinarios de IO proporciona beneficios mucho más modestos que estas aplicaciones reconocidas, las cifras que aparecen en la columna de la derecha de la tabla 1.1 reflejan el gran efecto que pueden tener los estudios grandes y bien diseñados de esta disciplina.

1.4 ALGORITMOS Y PAQUETES DE IO

Una parte primordial de este libro es la presentación de los **algoritmos** —procedimientos sistemáticos de solución— más importantes de la IO para resolver cierto tipo de problemas. Algunos de estos

■ **TABLA 1.1 Aplicaciones de la investigación de operaciones que se describirán en los recuadros de aplicación**

Organización	Área de aplicación	Sección	Ahorros anuales
Federal Express	Planeación logística de envíos	1.3	No estimados
Continental Airlines	Reasignación de tripulaciones a vuelos cuando ocurren interrupciones en el itinerario	2.2	$40 millones
Swift & Company	Mejora del desempeño en ventas y manufactura	3.1	$12 millones
Memorial Sloan-Kettering Cancer Center	Diseño de terapia de radiación	3.4	$459 millones
United Airlines	Plan para los programas de trabajo de los empleados en aeropuertos y oficinas de reservación	3.4	$6 millones
Welch's	Optimización del uso y movimiento de materias primas	3.6	$150 000
Samsung Electronics	Reducción de tiempos de manufactura y niveles de inventario	4.3	Ganancias adicionales de $200 millones
Pacific Lumber Company	Gestión de ecosistemas forestales a largo plazo	6.7	$398 millones VPN
Procter & Gamble	Rediseño del sistema de producción y distribución	8.1	$200 millones
Canadian Pacific Railway	Plan de rutas para un tren de carga	9.3	$100 millones
United Airlines	Reasignación de aviones a vuelos cuando ocurren interrupciones	9.6	No estimados
Ejército de Estados Unidos	Planeación logística de la Operación Tormenta del Desierto	10.3	No estimados
Air New Zealand	Programación de tripulaciones en una aerolínea	11.2	$6.7 millones
Taco Bell	Planeación de los programas de trabajo de los empleados de restaurantes	11.5	$13 millones
Gestión de desperdicios	Desarrollo de un sistema de administración de rutas para la recolección y disposición de basura	11.7	$100 millones
Bank Hapoalim Group	Desarrollo de un sistema de apoyo a las decisiones de asesores en inversiones	12.1	Ganancias adicionales de $31 millones
Sears	Rutas y programación de vehículos para servicio y entregas a domicilio	13.2	$42 millones
Conoco-Phillips	Evaluación de proyectos de exploración petrolera	15.2	No estimados
Oficina de compensaciones a los trabajadores	Gestión de solicitudes de incapacidad y rehabilitación de alto riesgo	15.3	$4 millones
Westinghouse	Evaluación de proyectos de investigación y desarrollo	15.4	No estimados
Merrill Lynch	Administración del riesgo de liquidez de líneas de crédito revolventes	16.2	Liquidez adicional de $4 mil millones
PSA Peugeot Citröen	Guía para el proceso de diseño de plantas de ensamble de automóviles eficientes	16.8	Utilidades adicionales de $130 millones
KeyCorp	Mejora de la eficiencia de los cajeros de banco	17.6	$20 millones
General Motors	Mejora de la eficiencia de líneas de producción	17.9	$90 millones
Deere & Company	Administración de inventarios a lo largo de una cadena de suministro	18.5	Reducción de $1 000 millones en inventario
Time Inc.	Administración de canales de distribución de revistas	18.7	Utilidades adicionales de $3.5 millones
Bank One Corporation	Administración de líneas de crédito y tasas de interés de tarjetas de crédito	19.2	Utilidades adicionales de $75 millones
Merrill Lynch	Análisis de precios de provisión de servicios financieros	20.2	Ganancias adicionales de $50 millones
AT&T	Diseño y operación de centros de atención telefónica	20.5	Utilidades adicionales de $750 millones

algoritmos son muy eficientes y casi siempre se utilizan para solucionar problemas que incluyen cientos o miles de variables. Además, se presenta una introducción acerca de cómo funcionan y qué los hace tan eficientes. Más adelante, estos algoritmos se utilizarán para resolver diversos problemas en una computadora. El **OR Courseware** que contiene el sitio web del libro (www.mhhe.com/hillier) es la herramienta clave para hacerlo.

Una característica especial del OR Courseware es el programa llamado **OR Tutor**, cuyo objetivo es ser una guía personal para ayudar en el aprendizaje de los algoritmos. Este programa contiene muchos *ejemplos de demostración* en los que se despliegan y explican los algoritmos en acción. Estas demostraciones complementan los ejemplos del libro.

Recuadro de aplicación

Federal Express (FedEx) es la compañía de transporte de paquetes más grande del mundo. Cada día de trabajo entrega más de 6.5 millones de documentos, paquetes y otros artículos a través de Estados Unidos y más de 220 países y territorios alrededor del mundo. En algunos casos, la entrega de estos embarques está garantizada para el día siguiente a las 10:30 a.m.

Los desafíos logísticos que implica proporcionar este servicio son extraordinarios. Los millones de embarques diarios deben ordenarse y ponerse en ruta de manera individual hacia la ubicación general correcta (por lo general por avión) para después ser entregados en el destino exacto (normalmente con un vehículo motorizado) en un tiempo sorprendentemente breve. ¿Cómo es esto posible?

La investigación de operaciones (IO) es el motor tecnológico que impulsa a esta compañía. Desde su fundación en 1973, la IO le ha ayudado a tomar sus decisiones de negocios más importantes, entre ellas inversiones en equipo, estructura de rutas, programación, finanzas y ubicación de instalaciones. Después de que la IO recibió el crédito por salvar a la compañía durante sus primeros años, se volvió una costumbre tener una representación de IO en las reuniones de la alta dirección y, en realidad, algunos de los vicepresidentes corporativos han surgido del destacable grupo de IO de FedEx.

FedEx ha sido reconocida como una compañía de clase mundial. De manera rutinaria se encuentra entre las mejores compañías de la lista anual de *Fortune Magazine* de "Las Compañías más Admiradas del Mundo". También fue la primera ganadora del prestigioso premio que ahora se conoce como INFORMS, el cual se otorga todos los años por la eficaz y repetida integración de la IO en la toma de decisiones de la organización en formas originales, variadas, novedosas y duraderas.

Fuente: R. O. Mason, J. L. McKenney, W. Carlson y D. Copeland, "Absolutely, Positively Operations Research: The Federal Express Story", *Interfaces*, **27**(2): 17-36, marzo-abril de 1997. (En nuestro sitio web www.mhhe.com/hillier se proporciona un vínculo con este artículo.)

Además, el OR Courseware incluye un paquete especial llamado **Interactive Operations Research Tutorial,** o **IOR Tutorial**. Este paquete innovador fue implementado en Java y está diseñado para mejorar la experiencia de aprendizaje de los estudiantes que utilicen este libro. El IOR Tutorial incluye muchas *rutinas interactivas* para ejecutar los algoritmos de manera dinámica y en un formato conveniente. La computadora realiza todos los cálculos de rutina mientras el estudiante centra su atención en aprender y ejecutar la lógica del algoritmo. Estas rutinas interactivas son una manera eficiente e ilustrativa de resolver muchos de los problemas de tarea. El IOR Tutorial también incluye otras herramientas útiles, como algunos *procedimientos automáticos* para ejecutar algoritmos y varios otros que ofrecen un despliegue gráfico de la forma en que la solución que proporciona un algoritmo varía a medida que cambian los datos del problema.

En la práctica, los algoritmos son ejecutados en paquetes de software comercial; por ello, es importante familiarizar al estudiante con la naturaleza de los programas que utilizará en la vida profesional. El OR Courseware incluye una gran cantidad de material para introducir los tres paquetes de mayor uso, descritos a continuación. Juntos, estos paquetes permiten resolver con gran eficiencia casi todos los modelos de IO que se presentan en este libro. Además, se agregan ciertas *rutinas automáticas* propias del OR Courseware sólo para algunos casos en los que estos paquetes no son aplicables.

En la actualidad, es común el uso del paquete de hojas de cálculo líder, *Microsoft Excel*, para elaborar pequeños modelos de IO en este formato. Después, se utiliza el **Excel Solver** para resolver los modelos, en ocasiones, en una versión mejorada, como el **Premium Solver for Education** incluido en el OR Courseware. Éste incluye archivos individuales de Excel, basados en el relativamente nuevo Excel 2007, para casi cada capítulo del libro. Cada vez que se presenta un ejemplo que pueda ser resuelto con Excel, se proporciona la formulación completa en una hoja de cálculo y se da la solución en el archivo de Excel del capítulo. En el caso de muchos modelos que aparecen en el libro se dispone de una *plantilla de Excel* que incluye las ecuaciones necesarias para resolver el modelo. Algunos *complementos de Excel* también se encuentran en el sitio web de este libro.

Después de muchos años, **LINDO** —y su lenguaje de modelado **LINGO**— aún es uno de los programas de software más populares para resolver modelos de investigación de operaciones. En la actualidad es posible bajar gratis de internet las versiones para estudiante, pero también fue incluido en el OR Courseware. En cuanto a Excel, cada vez que un ejemplo pueda ser resuelto con este paquete, se darán todos los detalles en un archivo de LINGO/LINDO para ese capítulo en el OR Courseware.

CPLEX es un software muy usado para resolver problemas grandes que son un reto en investigación de operaciones. Cuando se los debe enfrentar, también es común usar un *sistema de modelado* para elaborar el modelo matemático de manera eficiente e introducirlo en la computadora. **MPL** es

un sistema de modelado amigable que utiliza principalmente CPLEX para resolver los modelos. También utiliza otros paquetes que incluyen LINDO, CoinMP (que se explica en la sección 4.8), CONOPT (que se presenta en la sección 12.9), LGO (que se introduce en la sección 12.10) y BendX (útil para resolver algunos modelos estocásticos). Una versión para el estudiante de MPL, junto con las versiones para el estudiante más recientes de CPLEX y sus restantes paquetes de solución pueden obtenerse de manera gratuita en internet. Para conveniencia del lector, en el OR Courseware también se puede encontrar esta versión para el estudiante (que incluye todos los paquetes de solución mencionados). De nuevo, todos los ejemplos que puedan resolverse con estos paquetes se detallan en los archivos MPL/CPLEX de los capítulos correspondientes en el OR Courseware.

Estos tres paquetes y la manera de usarlos son descritos con más detalle en especial cerca del final de los capítulos 3 y 4. El apéndice 1 también proporciona documentación del OR Courseware y se incluye el OR Tutor y el IOR Tutorial.

Como una indicación acerca del material relevante que se incluye en el OR Courseware, al final de cada capítulo —a partir del tercero— aparecerá una lista de ayuda para el aprendizaje de este capítulo en nuestro sitio web. Como se explica al principio de la sección de problemas de cada capítulo, fueron colocados algunos símbolos a la izquierda del número del problema o del inciso cuando este material sea útil, y se incluyeron los ejemplos de demostración y las rutinas interactivas.

Otra ayuda para el aprendizaje en nuestro sitio web es un conjunto de ejemplos desarrollados **(Worked Examples)** para cada capítulo (del 3 en adelante). Estos ejemplos resueltos sirven de complemento a los del libro para que se utilicen cuando sea conveniente, sin interrumpir el flujo de material en las múltiples ocasiones en las que no es necesario un ejemplo adicional. Estos ejemplos complementarios también pueden ser útiles durante la preparación de un examen. Siempre que un ejemplo complementario de un tema en particular esté incluido en la sección de Worked Examples del sitio web, se lo mencionará en el texto del libro. Para asegurar que esta mención no sea pasada por alto, cada vez se marcarán las palabras **ejemplo adicional** (o algo similar) en negritas.

El sitio web también incluye un glosario de cada capítulo.

REFERENCIAS SELECCIONADAS

1. Bell, P. C., C. K. Anderson y S. P. Kaiser, "Strategic Operations Research and the Edelman Prize Finalist Applications 1989-1998", en *Operations Research,* **51**(1): 17-31, enero-febrero de 2003.
2. Gass, S. I. y A. A. Assad: *An Annotated Timeline of Operations Research: An Informal History,* Kluwer Academic Publishers (ahora Springer), Boston, 2005.
3. Gass, S. I. y C. M. Harris (eds.), *Encyclopedia of Operations Research and Management Science,* 2a. ed., Kluwer Academic Publishers, Boston, 2001.
4. Horner, P.: "History in the Making", en *OR/MS Today*, **29**(5): 30-39, octubre de 2002.
5. Horner, P. (ed.), "Special Issue: Executive's Guide to Operations Research", en *OR/MS Today,* Institute for Operations Research and the Management Sciences, **27**(3), junio de 2000.
6. Kirby, M. W.: "Operations Research Trajectories: The Anglo-American Experience from the 1940s to the 1990s", en *Operations Research,* **48**(5): 661-670, septiembre-octubre de 2000.
7. Miser, H. J., "The Easy Chair: What OR/MS Workers Should know About the Early Formative Years of Their Profession", en *Interfaces*, **30**(2): 99-111, marzo-abril de 2000.
8. Wein, L.M. (ed.), "50th Anniversary Issue", en *Operations Research* (un estudio especial que describe explicaciones personales de algunos desarrollos teóricos y prácticos clave recientes), **50**(1), enero-febrero de 2002.

PROBLEMAS

1.3-1. Seleccione una de las aplicaciones de investigación de operaciones que se mencionan en la tabla 1.1. Lea el artículo al que se hace referencia en el Recuadro de aplicación que se presenta en la sección que se muestra en la tercera columna. (En nuestro sitio web www.mhhe.com/hillier se proporciona un vínculo a todos estos artículos.) Escriba un resumen de dos páginas acerca de la aplicación y los beneficios que proporcionó; incluya los beneficios no financieros.

1.3-2. Seleccione tres de las aplicaciones de investigación de operaciones que se mencionan en la tabla 1.1. Para cada una, lea el artículo al que se hace referencia en el Recuadro de aplicación que se presenta en la sección que se muestra en la tercera columna. (En nuestro sitio web www.mhhe.com/hillier se proporciona un vínculo a todos estos artículos.) En cada caso, escriba un resumen de una página acerca de la aplicación y sus beneficios; incluya los no financieros.

1.3-3. Lea el artículo de referencia que describe todo el estudio de IO que se resume en el Recuadro de aplicación de la sección 1.3. Enliste los diferentes beneficios financieros y no financieros que resultaron de este estudio.

CAPÍTULO 2

Panorama del enfoque de modelado en investigación de operaciones

La mayor parte de este libro está dedicada a los métodos matemáticos de investigación de operaciones (IO). Esta situación resulta apropiada puesto que las técnicas cuantitativas constituyen la parte principal de lo que se conoce sobre el tema. Sin embargo, ello no significa que los estudios prácticos de IO sean, en esencia, ejercicios de matemáticas. Con frecuencia, el análisis matemático sólo representa una pequeña parte del trabajo. El propósito de este capítulo es dar a las cosas una mejor dimensión mediante la descripción de las etapas más importantes de un estudio característico de IO.

Una manera de resumir las fases usuales —traslapadas— de un estudio de investigación de operaciones es la siguiente:

1. Definición del problema de interés y recolección de datos relevantes.
2. Formulación de un modelo matemático que represente el problema.
3. Desarrollo de un procedimiento basado en computadora para derivar una solución para el problema a partir del modelo.
4. Prueba del modelo y mejoramiento de acuerdo con las necesidades.
5. Preparación para la aplicación del modelo prescrito por la administración.
6. Implementación.

En las siguientes secciones se analizará cada una de estas etapas.

Las referencias que se presentan al final del capítulo incluyen algunos estudios de IO, ganadores de premios, que proporcionan ejemplos excelentes de cómo ejecutar de manera correcta estas fases. A lo largo del capítulo se incorporarán algunos extractos de estos ejemplos. Si el lector decide aprender más acerca de estas aplicaciones, en el sitio web del libro www.mhhe.com/hillier se incluye un vínculo con los artículos que describen a detalle estos estudios de IO.

2.1 DEFINICIÓN DEL PROBLEMA Y RECOLECCIÓN DE DATOS

En contraste con los ejemplos de los libros de texto, la mayor parte de los problemas prácticos que enfrenta un equipo de IO son descritos, al principio, de una manera vaga e imprecisa. Por consiguiente, la primera actividad será el estudio del sistema relevante y el desarrollo de un resumen bien definido del problema que será analizado. Esta etapa incluye la determinación de los objetivos apropiados, las restricciones sobre lo que es posible hacer, las interrelaciones del área en estudio con otras áreas de la organización, los diferentes cursos de acción posibles, los límites de tiempo para tomar una decisión, etc. Este proceso de definición del problema es crucial, pues afectará de forma significativa la relevancia de las conclusiones del estudio. ¡Es difícil obtener una respuesta "correcta" a partir de un problema enfocado de manera "equivocada"!

Lo primero que debe reconocerse es que un equipo de IO, por lo general, trabaja a *nivel de asesoría.* A los miembros del equipo no se les presenta un problema y se les dice que lo resuelvan

como puedan, sino que asesoran a la administración (casi siempre un tomador de decisiones clave). El equipo realiza un análisis técnico detallado y después presenta recomendaciones. Este informe identifica cierto número de opciones atractivas, en particular con diferentes supuestos o para un rango diferente de valores, de algún parámetro que marca una política que puede ser evaluada sólo por esa administración: por ejemplo, la decisión entre *costo* y *beneficio.* La administración evalúa el estudio y sus recomendaciones, analiza una variedad de factores intangibles y toma una decisión final con base en su mejor juicio. Es vital que el equipo de IO tenga una visión al mismo nivel que la administración, incluso para identificar el problema "correcto" desde el punto de vista gerencial y que, a su vez, la administración le brinde apoyo sobre cualquier curso que tome el estudio.

Un aspecto muy importante de la formulación del problema es la determinación de los *objetivos apropiados.* Para hacerlo es necesario, en primer lugar, identificar a las personas de la administración que en realidad tomarán las decisiones concernientes al sistema en estudio, y después escudriñar el pensamiento de estos individuos en relación con los objetivos pertinentes. (La inclusión del tomador de decisiones desde el principio es esencial para obtener su apoyo durante la realización del estudio.)

Por su naturaleza, a la IO le concierne el bienestar de *toda la organización*, no sólo de algunos componentes. Un estudio de IO trata de encontrar soluciones óptimas globales, y no soluciones subóptimas aunque sean lo mejor para uno de los componentes. Desde un punto de vista ideal, los objetivos formulados deben coincidir con los de toda la organización; sin embargo, esta coincidencia no siempre es conveniente. Muchos problemas interesan sólo a una parte de la organización, de manera que el análisis sería demasiado extenso si los objetivos fueran generales y se prestara atención especial a todos los efectos secundarios sobre el resto de la organización. En lugar de ello, los objetivos de un estudio deben ser tan específicos como sea posible, siempre y cuando consideren las metas principales del tomador de decisiones y mantengan un nivel razonable de congruencia con los objetivos de niveles más elevados.

Cuando se trata de organizaciones lucrativas, un enfoque posible para no caer en un problema de suboptimización es utilizar la *maximización de la ganancia a largo plazo*, considerando el valor del dinero en el tiempo como un objetivo único. El adjetivo *a largo plazo* indica que este objetivo proporciona la flexibilidad necesaria para considerar actividades que no se traducen *de inmediato* en ganancias, como los proyectos de investigación y desarrollo, pero que deberán hacerlo *con el tiempo* para que valgan la pena. Este enfoque tiene muchas ventajas. El objetivo es tan específico como para usarlo en forma adecuada y al mismo tiempo lo bastante amplio como para tomar en cuenta la meta básica de las organizaciones lucrativas. En realidad, algunas personas piensan que cualquier otro objetivo legítimo se puede traducir en ganancias.

Sin embargo, en la práctica, muchas organizaciones lucrativas no utilizan este enfoque. Algunos estudios de corporaciones estadounidenses han demostrado que la administración tiende a adoptar la meta de *ganancias satisfactorias* combinada con *otros objetivos*, en lugar de enfocarse en la maximización de la ganancia a largo plazo. Algunos de estos *otros* objetivos pueden ser conservar la estabilidad de las ganancias, aumentar o conservar la participación de mercado con que se cuenta, permitir la diversificación de productos, mantener precios estables, mejorar las condiciones y el ánimo de los trabajadores, mantener el control familiar sobre el negocio o incrementar el prestigio de la compañía. Si se satisfacen estos objetivos, tal vez se logre maximizar las ganancias a largo plazo, pero la relación puede ser tan oscura que quizá sea mejor no incorporarlos.

Existen otras consideraciones que incluyen responsabilidades sociales muy distintas al objetivo de las ganancias. Las cinco partes que son afectadas por una empresa de negocios que se localizan en un país determinado son: 1) los *dueños* (accionistas, etc.), que desean obtener ganancias (dividendos, valuación de acciones, etc.); 2) los *empleados*, que aspiran a un empleo seguro con un salario razonable; 3) los *clientes*, que quieren un producto confiable a un precio justo; 4) los *proveedores*, que desean integridad y un precio de venta razonable para sus bienes, y 5) el *gobierno* y, por ende, la *nación*, que quiere el pago de impuestos justo y que se tome en cuenta el interés común. Las cinco partes hacen contribuciones esenciales a la empresa; ésta no debe servir a ninguna de ellas para explotar a las otras. De la misma manera, las corporaciones internacionales adquieren las obligaciones adicionales de cumplir con una práctica social responsable. En consecuencia, aunque se acepte que obtener ganancias es la responsabilidad primordial de la administración, lo cual, en última instancia, beneficia a las cinco partes, también deben reconocerse estas responsabilidades sociales más extensas.

Es común que los equipos de IO pasen mucho tiempo en la *recolección de los datos relevantes* del problema. Se necesitan muchos datos para lograr la comprensión exacta del problema y así proporcionar el insumo adecuado para el modelo matemático que se elaborará en la siguiente etapa del estudio. Con frecuencia, al inicio de éste no se dispone de muchos datos necesarios, ya sea porque nunca se guardó la información o porque lo que se guardó cayó en la obsolescencia o se almacenó en una forma incorrecta. En consecuencia, muchas veces se debe instalar un nuevo *sistema de información general* para reunir los datos sobre la marcha y en la forma adecuada. El equipo de IO debe destinar un tiempo considerable para recabar la ayuda de otros miembros clave de la organización, entre ellos especialistas en *tecnología de la información* (TI), para rastrear todos los datos vitales. Aun con este esfuerzo, muchos datos pueden ser blandos, es decir, estimaciones burdas basadas sólo en juicios personales. A menudo, el equipo de IO debe utilizar una gran cantidad de tiempo para mejorar la precisión de los datos y al final tendrá que trabajar con lo mejor que pudo obtener.

Debido a la expansión del uso de bases de datos y el crecimiento explosivo de su tamaño en los años recientes, en la actualidad los equipos de IO a menudo se encuentran con que su problema más grande con los datos es que existen demasiados. Puede haber miles de fuentes de información, por lo cual la cantidad total de datos debe medirse en gigabytes o incluso en terabytes. En este entorno, la localización de los datos relevantes y la identificación de patrones interesantes pueden convertirse en tareas abrumadoras. Una de las herramientas más modernas de los equipos de IO que aborda este problema es una técnica denominada **extracción de datos**. Los métodos para aplicarla tratan de descubrir patrones interesantes dentro de las grandes fuentes de información que puedan conducir a una toma de decisiones útiles. (La segunda referencia seleccionada al final del capítulo proporciona una base más sólida acerca de la extracción de datos.)

Ejemplo. A finales de la década de los noventa, las compañías de servicios financieros generales sufrieron el ataque de las firmas de correduría electrónica que ofrecían costos de compraventa financiera muy bajos. **Merrill Lynch** respondió con la realización de un gran estudio de IO que recomendó la revisión completa de la manera en que cobraba sus servicios, desde una opción basada en activos de servicio completo —cargo de un porcentaje fijo del valor de los activos en vez de hacerlo por transferencias individuales— hasta una opción de bajo costo para los clientes que deseaban invertir en línea de manera directa. La *recolección y el procesamiento de datos* tuvieron un papel fundamental en el estudio. Para analizar el efecto del comportamiento de cada uno de los clientes en respuesta a diferentes opciones, el equipo decidió montar una base de datos de clientes con una capacidad de 200 gigabytes, la cual debía contener 5 millones de clientes, 10 millones de cuentas, 100 millones de registros de transacciones y 250 millones de registros contables. Este objetivo requirió combinar, reconciliar, filtrar y limpiar datos procedentes de muchas bases de datos. La adopción de las recomendaciones del estudio produjo un incremento de cerca de 50 000 millones de dólares en la posesión de activos de sus clientes y casi 80 millones de dólares en ganancias adicionales. (La referencia seleccionada A2 describe este estudio a detalle.)

2.2 FORMULACIÓN DE UN MODELO MATEMÁTICO

Una vez que el tomador de decisiones define el problema, la siguiente etapa consiste en reformularlo de manera conveniente para su análisis. La forma convencional en que la investigación de operaciones logra este objetivo es mediante la construcción de un modelo matemático que represente la esencia del problema. Antes de analizar cómo se elaboran los modelos de este tipo se explorará su naturaleza general y, en particular, la de los modelos matemáticos.

Los modelos, o representaciones idealizadas, son una parte integral de la vida diaria. Entre los ejemplos más comunes pueden citarse modelos de avión, retratos, globos terráqueos y otros. De igual manera, los modelos tienen un papel importante en la ciencia y los negocios, como lo hacen patente los modelos del átomo y de las estructuras genéticas, las ecuaciones matemáticas que describen las leyes físicas del movimiento o las reacciones químicas, las gráficas, los organigramas y los sistemas contables en la industria. Esos modelos son invaluables, pues extraen la esencia del material de estudio, muestran sus interrelaciones y facilitan el análisis.

Los modelos matemáticos también son representaciones idealizadas, pero están expresados en términos de símbolos y expresiones matemáticas. Las leyes de la física como $F = ma$ y $E = mc^2$ son ejemplos familiares. En forma parecida, el modelo matemático de un problema industrial está conformado por el sistema de ecuaciones y expresiones matemáticas relacionadas que describen la esencia del problema. De esta forma, si deben tomarse n decisiones cuantificables relacionadas entre sí, se representan como **variables de decisión** $(x_1, x_2, \ldots, x_n)$ para las que se deben determinar los valores respectivos. En consecuencia, la medida de desempeño adecuada (por ejemplo, la ganancia) se expresa como una función matemática de estas variables de decisión (por ejemplo, $P = 3x_1 + 2x_2 + \ldots + 5x_n$). Esta función se llama **función objetivo**. También se expresan en términos matemáticos todas las limitaciones que se puedan imponer sobre los valores de las variables de decisión, casi siempre en forma de ecuaciones o desigualdades (como $x_1 + 3x_1 x_2 + 2x_2 \leq 10$). Con frecuencia, tales expresiones matemáticas de las limitaciones reciben el nombre de **restricciones**. Las constantes (los coeficientes o el lado derecho de las expresiones) de las restricciones y de la función objetivo se llaman **parámetros** del modelo. El modelo matemático puede decir entonces que el problema es elegir los valores de las variables de decisión de manera que se maximice la función objetivo, sujeta a las restricciones dadas. Un modelo de este tipo, y algunas de sus variantes menores, tipifican los modelos que se analizan en investigación de operaciones.

La determinación de los valores apropiados que deben asignarse a los parámetros del modelo —un valor por parámetro— es una tarea crítica y a la vez un reto en el proceso de construcción del modelo. Al contrario de los problemas que se presentan en los libros donde se proporcionan estos números, la determinación de los valores de los parámetros en los problemas reales requiere la *recolección de los datos relevantes.* Como se vio en la sección anterior, a menudo la recolección de datos exactos es difícil. Por lo tanto, es común que el valor asignado a un parámetro sea, por necesidad, sólo una estimación. Debido a la incertidumbre sobre el valor real del parámetro es importante analizar la forma de cómo cambiaría —si lo hace— la solución derivada del problema cuando el valor asignado al parámetro cambia por otros valores posibles. Este proceso, que se conoce como **análisis de sensibilidad**, se estudiará en la siguiente sección (y en gran parte del capítulo 6).

Aun cuando se hable de "el" modelo matemático de un problema industrial, por lo general los problemas reales no pueden ser representados por un solo modelo "correcto". En la sección 2.4 se describe la manera en que el proceso de prueba de un modelo conduce a una serie de modelos que proporcionan representaciones cada vez mejores del problema real. Incluso, es posible desarrollar dos o más tipos de modelos diferentes para analizar el mismo problema.

A lo largo de este libro se proporcionarán numerosos ejemplos de modelos matemáticos. En los capítulos siguientes se estudia cierta clase de modelo con una importancia especial, denominado **modelo de programación lineal**, en el que las funciones matemáticas que aparecen tanto en la función objetivo como en las restricciones, son funciones lineales. En el capítulo 3 se construyen modelos específicos de programación lineal que se ajustan a diversos tipos de problemas, tales como determinar 1) la mezcla de productos que maximiza la ganancia, 2) el diseño de la terapia de radiación que combata de manera eficaz un tumor y que al mismo tiempo minimice el daño al tejido sano circundante, 3) la asignación de hectáreas a distintos cultivos para maximizar el rendimiento total neto y 4) la combinación de métodos de control de contaminación que logre los estándares de calidad del aire a un costo mínimo.

Los modelos matemáticos tienen muchas ventajas sobre una descripción verbal del problema. La más obvia es que describe un problema en forma mucho más concisa. Esta característica tiende a hacer más comprensible toda la estructura del problema y ayuda a revelar las relaciones importantes causa-efecto. En segundo lugar, indica con mayor claridad qué datos adicionales son importantes para el análisis. También facilita el manejo total del problema y, al mismo tiempo, el estudio de sus interrelaciones. Por último, un modelo matemático forma un puente para el empleo de técnicas matemáticas y computadoras de alto poder para analizar el problema. Sin duda, existe una amplia disponibilidad de paquetes de software para resolver muchos tipos de modelos matemáticos, en computadoras personales y de gran poder.

Por otro lado, existen obstáculos que se deben evitar cuando se utilizan modelos matemáticos. Un modelo es, por necesidad, una idealización abstracta del problema, por lo que casi siempre se requieren aproximaciones y supuestos de simplificación si se desea que el modelo sea *manejable* (susceptible de ser resuelto). Por lo tanto, debe tenerse cuidado de que el modelo sea siempre una representación válida del problema. El criterio adecuado para juzgar la validez de un modelo es si

predice o no con suficiente exactitud los efectos relativos de los diferentes cursos de acción, para poder tomar una decisión que tenga sentido. No es necesario incluir detalles sin importancia o factores que tienen casi el mismo efecto sobre todas las opciones. Ni siquiera es necesario que la magnitud absoluta de la medida de desempeño sea casi correcta para las diferentes alternativas, siempre que sus valores relativos —es decir, las diferencias entre sus valores— sean bastante precisos. En consecuencia, todo lo que se requiere es que exista una alta *correlación* entre la predicción del modelo y lo que ocurre en la vida real. Para asegurar que este requisito se cumpla, es importante hacer un número considerable de *pruebas* del modelo y las modificaciones consecuentes, que serán el tema de la sección 2.4. Aunque en el orden del libro esta fase de pruebas se haya colocado después, gran parte del trabajo de *validación del modelo* se lleva a cabo en la etapa de construcción para que sirva de guía para elaborar el modelo matemático.

En la etapa de desarrollo del modelo se recomienda empezar con una versión muy sencilla y avanzar de manera evolutiva hacia paradigmas más elaborados que reflejen mejor la complejidad del problema real. Este proceso de *enriquecimiento del modelo* continúa sólo mientras sea manejable. La decisión básica que debe tomarse oscila entre la *precisión* y el *manejo* del modelo. (Vea en la referencia 8 una descripción detallada de este proceso.)

Un paso crucial en la formulación de un modelo de IO es la construcción de la función objetivo. Esta tarea requiere desarrollar una medida cuantitativa del desempeño asociado a cada objetivo que el tomador de decisiones identifica cuando define el problema. Si en el estudio se contempla más de un objetivo, es necesario transformar y combinar las medidas respectivas en una medida compuesta de eficacia llamada **medida global de desempeño**. Esta medida compuesta puede ser algo tangible —ganancias— y corresponder a una meta más alta de la organización, o puede ser abstracta —utilidad—. En este caso, desarrollar una función de utilidad puede ser complejo y requerir una comparación cuidadosa de los objetivos y su importancia relativa. Una vez desarrollada la medida global de desempeño, la función objetivo expresa esta medida como una función matemática de las variables de decisión. De manera alternativa, existen métodos que contemplan al mismo tiempo y en forma explícita objetivos múltiples; en el capítulo 7 se analiza uno de ellos (programación de metas).

Ejemplo. La oficina responsable de control del agua y los servicios públicos del gobierno de Holanda, el **Rijkswaterstaat**, contrató un importante estudio de IO para guiar el desarrollo de una nueva política de administración del vital líquido. La nueva política ahorró cientos de millones de dólares en gastos de inversión y redujo el daño agrícola en alrededor de 15 millones de dólares anuales, al mismo tiempo que disminuyó la contaminación térmica y la debida a las algas. En lugar de elaborar sólo *un* modelo matemático se desarrolló un sistema integrado y comprensivo de ¡50 modelos! Más aún, en el caso de algunos modelos se desarrollaron versiones sencillas y complejas. La versión sencilla se usó para adquirir una visión básica que incluyó el análisis de intercambios. La versión compleja se utilizó después, en las corridas finales del análisis o cuando se deseaba mayor exactitud o más detalle en los resultados. El estudio completo de IO involucró de manera directa a más de 125 personas-año de esfuerzo, más de un tercio de ellas en la recolección de datos, creó varias docenas de programas de computadora y estructuró una enorme cantidad de datos. (La referencia seleccionada A7 describe el estudio a detalle.)

2.3 OBTENCIÓN DE SOLUCIONES A PARTIR DEL MODELO

Una vez formulado el modelo matemático del problema en estudio, la siguiente etapa de un trabajo de IO consiste en desarrollar un procedimiento, por lo general en computadora, para obtener una solución a partir de este modelo. Puede pensarse que ésta debe ser la parte principal del estudio, pero, en realidad, en la mayoría de los casos no lo es. De hecho, a veces ésta es una etapa bastante sencilla, en la que se aplica uno de los dos **algoritmos** —procedimientos iterativos de solución— de investigación de operaciones en una computadora mediante el uso de algunos de los paquetes de software disponibles. Para el investigador de operaciones experimentado, encontrar la solución es la parte divertida, mientras que el verdadero trabajo se encuentra en las etapas anteriores y en las subsecuentes, entre las que se incluyen el *análisis posóptimo*, tema que se explicará más adelante en esta sección.

Recuadro de aplicación

Continental Airlines es una importante compañía estadounidense que transporta pasajeros, carga y correo. Opera más de 2 000 salidas diarias a muchos más de 100 destinos nacionales y casi 100 internacionales.

Las aerolíneas como Continental enfrentan a diario interrupciones en el itinerario debido a eventos inesperados, entre los que se destacan inclemencias del clima, problemas mecánicos en los aviones e indisponibilidad de las tripulaciones. Estas interrupciones pueden ocasionar retrasos y cancelaciones en los vuelos. En consecuencia, las tripulaciones pueden no estar en posición de prestar el servicio en sus vuelos programados restantes. Las líneas aéreas deben reasignar tripulaciones con rapidez para cubrir los vuelos abiertos y para regresarlos a sus programas originales de una manera eficiente en cuanto a costos al mismo tiempo que cumple todas las regulaciones gubernamentales, obligaciones contractuales y elevados estándares de calidad de vida.

Para enfrentar estos problemas, un equipo de IO en Continental Airlines desarrolló un *modelo matemático* detallado para reasignar tripulaciones a vuelos tan pronto como surgían las emergencias. Como la aerolínea tiene miles de tripulaciones y vuelos diarios, el modelo necesitaba ser enorme para poder considerar todas las posibles asignaciones de tripulaciones a vuelos. Por lo tanto, el modelo tiene *millones de variables de decisión* y *muchos miles de restricciones*. En su primer año de uso (principalmente en 2001), el modelo se aplicó cuatro veces para recuperarse de interrupciones de itinerario muy importantes (dos tormentas de nieve, una inundación y los ataques terroristas del 11 de septiembre). Ello produjo *ahorros de aproximadamente* **40 millones de dólares**. Las aplicaciones subsiguientes se extendieron a muchas interrupciones menores diarias.

Aunque después otras aerolíneas comenzaron a aplicar investigación de operaciones de una manera similar, esta ventaja inicial sobre las demás compañías para recuperarse con más rapidez de interrupciones de itinerario con menos vuelos retrasados y cancelados colocó a Continental Airlines en una posición fuerte cuando la industria enfrentó un periodo difícil durante los años iniciales del siglo XXI. Esta iniciativa permitió a la empresa ganar en 2002 el primer lugar en la competencia internacional por el Premio Franz Edelman al desempeño en investigación de operaciones y ciencias de la administración.

Fuente: G. Yu, M. Argüello, C. Song, S. M. McGowan y A. White, "A New Era for Crew Recovery at Continental Airlines", en *Interfaces*, **33**(1): 5-22, enero-febrero, 2003. (En nuestro sitio web se proporciona un vínculo con este artículo, www.mhhe.com/hillier.)

Como la mayor parte de este libro está dedicada a la obtención de soluciones para los distintos e importantes tipos de modelos matemáticos, en este momento no es necesario aclarar nada al respecto. Sin embargo, sí lo es presentar la naturaleza de ellas.

Un tema común en IO es la búsqueda de una **solución óptima**, es decir, la mejor. Sin duda, como aquí se presenta, se han desarrollado muchos procedimientos para encontrarla en cierto tipo de problemas, pero es necesario reconocer que estas soluciones son óptimas sólo respecto del modelo elaborado. Además, como éste, por necesidad, es una idealización más que una representación exacta del problema real, no existe una garantía utópica de que sea la mejor solución que pueda implantarse. Existen demasiados imponderables e incertidumbres asociados con los problemas reales; sin embargo, si el modelo está bien formulado y verificado, la solución debe tender a constituirse en una buena aproximación de un curso de acción ideal en la realidad. Por esto, más que empeñarse en exigir lo imposible, la prueba del éxito de un estudio de IO debe ser si proporciona o no una mejor guía para la toma de decisiones que la que se puede obtener por otros medios.

Herbert Simon, eminente científico de la administración y premio Nobel de economía, introdujo el concepto de que en la práctica es mucho más frecuente **satisfizar** que optimizar. Al inventar el término *satisfizar* como una combinación de *satisfacer* y *optimizar*, Simon describe la tendencia de los administradores a buscar una solución que sea "lo suficientemente buena" para el problema que se enfrenta. En lugar de intentar el desarrollo de una medida global de desempeño para conciliar de manera óptima los conflictos entre los diferentes objetivos deseables, incluso los criterios bien establecidos para juzgar el desempeño de los distintos segmentos de la organización, es posible utilizar un enfoque más pragmático. Las metas se pueden establecer de manera que marquen los niveles mínimos satisfactorios de desempeño en las diferentes áreas, con base quizá en niveles de desempeño anteriores o en los logros de la competencia. Si se encuentra una solución que permita que todas estas metas sean cumplidas, es posible que sea adoptada sin más requisitos. Ésta es la naturaleza de satisfizar.

La distinción entre optimizar y "satisfizar" refleja la diferencia entre la teoría y la realidad, disparidad que con frecuencia se encuentra al tratar de implantar esa teoría en la práctica. En pa-

labras de uno de los líderes ingleses de la investigación de operaciones, Samuel Eilon, "optimizar es la ciencia de lo absoluto; satisfizar es el arte de lo factible".[1]

Los equipos de IO intentan incorporar al proceso de toma de decisiones la mayor cantidad posible de la "ciencia de lo absoluto". Sin embargo, un equipo que trabaja con éxito debe reconocer la necesidad más importante del tomador de decisiones: obtener una guía satisfactoria para sus acciones en un periodo razonable. Por lo tanto, la meta de un estudio de investigación de operaciones debe ser la realización del proceso de manera óptima, sin importar si implica una solución óptima para el modelo. Además de buscar la ciencia de lo absoluto, el equipo debe tomar en cuenta el costo del estudio y las desventajas de retrasar su terminación, y después, intentar maximizar los beneficios netos que resulten de dicho estudio. Al reconocer este concepto, en ocasiones los equipos de investigación de operaciones utilizan sólo **procedimientos heurísticos** —es decir, procedimientos de diseño intuitivo que no garantizan una solución óptima— para encontrar una buena **solución subóptima**. Esto ocurre con más frecuencia en los casos en que el tiempo o el costo para encontrar una solución óptima para un modelo adecuado del problema son muy grandes. En años recientes se han logrado grandes progresos en el desarrollo de procedimientos **metaheurísticos** eficientes y eficaces; estos procedimientos proporcionan una estructura general y directrices estratégicas para diseñar un procedimiento heurístico específico que se ajuste a un tipo particular de problema. El uso del enfoque metaheurístico (capítulo 13) continúa en crecimiento.

Hasta aquí ha quedado claro que un estudio de investigación de operaciones trata de encontrar una solución única, de la que puede o no requerirse que sea óptima. En realidad, esto casi nunca es lo que se busca. Una solución óptima para el modelo original puede estar muy alejada del ideal en el caso del problema real, de manera que es necesario hacer un análisis adicional. El **análisis posóptimo** —que se lleva a cabo después de encontrar una solución óptima— constituye una parte muy importante de la mayoría de los estudios de investigación de operaciones. Este análisis también se conoce como **análisis de qué pasa si**, puesto que involucra algunas preguntas acerca de *qué* pasaría con la solución óptima *si* se hubieran considerado supuestos diferentes sobre las condiciones futuras. Con frecuencia, los administradores que tomarán las últimas decisiones son quienes hacen estas preguntas y no el equipo de IO.

Con el advenimiento de los poderosos paquetes de software de hojas de cálculo, en la actualidad el papel central del análisis posóptimo suele tenerlo una de estas hojas. Una de sus grandes ventajas es la facilidad con que cualquier persona puede usarlas, incluidos los administradores, para ver qué pasa con una solución óptima cuando se introducen cambios en el modelo. Este proceso de experimentar con cambios en el modelo también puede ser útil para llegar a comprender el comportamiento del modelo y adquirir mayor confianza en su validez.

En parte, el análisis posóptimo implica llevar a cabo un **análisis de sensibilidad** para determinar qué parámetros del modelo son críticos —los "parámetros sensibles"— cuando se determina la solución. Una definición común de *parámetros sensibles* —que se usará aquí— es:

> En el caso de un modelo matemático cuyos parámetros tienen valores específicos, los **parámetros sensibles** del modelo son aquellos cuyos valores no se pueden cambiar sin que la solución óptima también cambie.

La identificación de los parámetros sensibles es importante, porque determina aquellos cuyos valores deben asignarse con más cuidado para evitar distorsiones en los resultados que generará el modelo.

Por lo general, el valor de un parámetro es una *estimación* de alguna cantidad —por ejemplo, la ganancia unitaria— cuyo valor exacto se conocerá sólo después de poner en práctica la solución. Por lo tanto, después de identificar los parámetros sensibles se deben realizar estimaciones más precisas y cuidadosas de cada uno de ellos, o por lo menos del intervalo de valores posibles. Después, se busca una solución aceptable para todas las combinaciones de valores posibles de los parámetros sensibles.

Si la solución se implanta sobre la marcha, cualquier cambio posterior en el valor de un parámetro sensible señala de inmediato la necesidad de cambiar la solución.

[1] S. Eilon, "Goals and Constraints in Decision-making", en *Operational Research Quarterly,* **23:** 3-15, 1972, conferencia anual de la Canadian Operational Research Society, 1971.

En algunos casos, ciertos parámetros del modelo representan políticas de decisión, como asignación de recursos. Si es así, con frecuencia existe flexibilidad sobre los valores dados a estos parámetros. Tal vez algunos se puedan aumentar si otros disminuyen. El análisis posóptimo incluye la investigación de estas compensaciones.

Junto con la etapa de estudio que se presenta en la siguiente sección: prueba del modelo, el análisis posóptimo incluye obtener un conjunto de soluciones que comprende una serie de aproximaciones, cada vez más precisas, al curso de acción ideal. De esta forma, las debilidades aparentes de la solución inicial se usan para sugerir mejoras al modelo, a sus datos de entrada y quizás al procedimiento de solución. De esta forma se obtiene una nueva solución, y el ciclo se repite. Este proceso sigue hasta que las mejoras a soluciones sucesivas son demasiado pequeñas como para justificar su reiteración. Aun entonces pueden presentarse a la administración varias soluciones posibles —quizá soluciones que son óptimas para una de varias versiones convincentes del modelo y sus datos de entrada— para hacer la selección final. Como se explicó en la sección 2.1, esta presentación de soluciones alternativas se lleva a cabo cuando la elección final entre ellas debe basarse en consideraciones que es mejor dejar al juicio de la administración.

Ejemplo. Considere de nuevo el estudio de IO del **Rijkswaterstaat** sobre la política nacional de administración de agua en Holanda, que se presentó al final de la sección anterior. Este estudio no concluyó con la recomendación de una sola solución. Más bien, se identificaron, analizaron y compararon varias alternativas atractivas. La elección final, que se dejó al proceso político del gobierno de Holanda, culminó con la aprobación del Parlamento. El *análisis de sensibilidad* tuvo un papel importante en este estudio. Por ejemplo, ciertos parámetros de los modelos representaron estándares ecológicos. El análisis de sensibilidad incluyó la evaluación del efecto en los problemas de agua si los valores de estos parámetros fueran cambiados de los estándares ecológicos actuales a otros valores razonables. Se usó también para evaluar el efecto de cambios en los supuestos de los modelos, por ejemplo, el supuesto sobre el efecto de tratados internacionales futuros en materia de contaminación que pudiera firmar Holanda. También se analizaron varios *escenarios*, como años secos o lluviosos en extremo, para lo cual se asignaron las probabilidades adecuadas.

2.4 PRUEBA DEL MODELO

En algunos aspectos, la elaboración de un modelo matemático grande es análogo al desarrollo de un programa de computadora grande. Cuando se completa la primera versión es inevitable que contenga muchas fallas. El programa debe ser probado de manera exhaustiva para tratar de encontrar y corregir tantas fallas como sea posible. Con el tiempo, después de una larga serie de programas mejorados, el programador —o equipo de programación— concluye que el programa actual proporciona, en general, resultados razonablemente válidos. Aunque sin duda quedarán algunas fallas ocultas —y quizá nunca sean detectadas— se habrá eliminado la cantidad suficiente de problemas mayores como para que su utilización sea confiable.

De manera similar, es inevitable que la primera versión de un modelo matemático complejo tenga muchas fallas. Sin duda, algunos factores o interrelaciones relevantes no fueron incorporados a él y algunos parámetros no fueron estimados con precisión. Estas circunstancias no se pueden eludir dadas las dificultades de comunicación y comprensión de todos los aspectos y sutilezas de un problema operacional complejo, así como la dificultad de recolectar datos confiables. Por lo tanto, antes de usar el modelo debe probarse de manera exhaustiva para intentar identificar y corregir la mayor cantidad posible de fallas. Con el tiempo, después de una larga serie de modelos mejorados, el equipo de investigación de operaciones concluye que el modelo actual produce resultados razonablemente válidos. Aunque sin duda quedarán algunos problemas menores ocultos en el modelo, las fallas importantes habrán sido eliminadas de manera que el uso del modelo sea confiable.

Este proceso de prueba y mejoramiento de un modelo para incrementar su validez se conoce como **validación del modelo**.

Es difícil describir cómo se lleva a cabo la validación del modelo porque el proceso depende en gran parte del problema bajo estudio y del modelo que se elaboró. Sin embargo, se harán algunos comentarios generales y después se darán algunos ejemplos (vea más detalles en la referencia 3).

Como el equipo de IO puede emplear meses en las actividades de desarrollo de todas las piezas detalladas del modelo, es fácil que “los árboles le impidan ver el bosque”. Después de completar los detalles (“los árboles”) de la versión inicial del modelo, una buena manera de comenzar las pruebas es observarlo en forma global (“el bosque”) para verificar los errores u omisiones obvias. El grupo que lleva a cabo esta revisión debe, de preferencia, incluir por lo menos a una persona que no haya participado en la elaboración del modelo. Cuando se examina de nuevo la formulación del problema y se la compara con el modelo pueden descubrirse algunos errores. También es útil asegurarse de que todas las expresiones matemáticas de las *dimensiones* de las unidades que se emplearon sean *congruentes.* Además, puede obtenerse un mejor conocimiento de la validez del modelo si se modifican los valores de los parámetros de entrada y/o de las variables de decisión, y se comprueba que los resultados del modelo se comportan de una manera plausible. A menudo estos resultados son reveladores, en especial cuando son asignados a los parámetros o variables valores extremos cercanos a sus máximos o a sus mínimos.

Un planteamiento más sistemático de la prueba del modelo se logra mediante el empleo de una **prueba retrospectiva**. Cuando es aplicable, esta prueba utiliza datos históricos y reconstruye el pasado para determinar si el modelo y la solución resultante hubieran tenido un buen desempeño, si se hubieran usado. La comparación de la eficacia de este desempeño hipotético con lo que en realidad ocurrió indica si la utilización del modelo tiende a generar mejoras significativas respecto de la práctica actual. Puede indicar también áreas en las que el modelo tiene fallas y requiere modificaciones. Aún más, cuando se emplean las alternativas de solución y se estiman sus desempeños históricos hipotéticos, se pueden reunir evidencias sobre la precisión del modelo para predecir los efectos relativos de los diferentes cursos de acción.

Por otra parte, la prueba retrospectiva tiene la desventaja de que se basa en los mismos datos que sirvieron para formular el modelo. Entonces, la pregunta crucial es si el pasado en realidad representa el futuro. Si no es así, el modelo puede tener un desempeño distinto en el futuro del que haya tenido en el pasado.

Para evitar esta desventaja de la prueba retrospectiva, a veces es útil continuar con las circunstancias actuales durante una temporada. Este recurso proporcionará nuevos datos con los cuales no se contaba cuando el modelo fue construido, los que se pueden emplear de la manera descrita para evaluar el modelo.

Es importante documentar el proceso que se utiliza para llevar a cabo las pruebas de la validación del modelo, pues ello ayuda a aumentar la confianza de los futuros usuarios del paradigma. Más aún, si en el futuro surgen preocupaciones sobre el modelo, esta documentación ayudará a diagnosticar en dónde pueden encontrarse los problemas.

Ejemplo. Considere un estudio de IO realizado para **IBM** cuyo objetivo era integrar su red nacional de inventarios de refacciones para mejorar el servicio a los clientes, al mismo tiempo que reducir el valor de aquéllos en más de 250 millones de dólares y ahorrar otros 20 millones de dólares anuales mediante el mejoramiento de la eficiencia operacional. Un aspecto en particular interesante de la etapa de validación del modelo en este estudio fue la manera en que fueron incorporados al proceso de prueba los *usuarios futuros* del sistema de inventarios. Debido a que estos usuarios futuros —los administradores de IBM de las áreas funcionales responsables de la implantación del sistema de inventarios —dudaban del sistema que se desarrollaba, fueron asignados representantes de la empresa a un *equipo de usuarios* que tendría la función de asesorar al equipo de IO. Una vez desarrollada la versión preliminar del nuevo sistema —basada en un sistema de inventarios de multiniveles— se llevó a cabo una *prueba preliminar de implementación.* La extensa retroalimentación por parte del equipo de usuarios generó mejoras importantes en el sistema propuesto. (La referencia seleccionada A5 describe este estudio a detalle.)

2.5 PREPARACIÓN PARA APLICAR EL MODELO

¿Qué pasa después de completar la etapa de pruebas y desarrollar un modelo aceptable? Si el modelo va a usarse varias veces, el siguiente paso es instalar un *sistema* bien documentado para aplicarlo según lo establecido por la administración. Este sistema debe incorporar el modelo y el procedimiento de solución —que incluye el análisis posóptimo— y los procedimientos operativos

para su implantación. Así, aun cuando cambie el personal, el sistema puede ser consultado de manera periódica para proporcionar una solución numérica específica.

Este sistema casi siempre *se diseña para computadora.* En realidad, con frecuencia se necesita un número considerable de programas integrados. Las *bases de datos* y los *sistemas de información administrativos* pueden proporcionar la entrada actualizada para el modelo cada vez que se use, en cuyo caso se necesitan programas de interfaz, es decir, interacción con el usuario. Después de aplicar un procedimiento de solución —otro programa— al modelo, puede ser que los programas adicionales manejen la implementación de los resultados de manera automática. En otros casos, se instala un sistema *interactivo* de computadora llamado **sistema de apoyo para las decisiones** para ayudar a la administración a usar los datos y modelos y así apoyar (no sustituir) su toma de decisiones. Otro programa puede generar *informes gerenciales* (en lenguaje administrativo) que interpreten la salida del modelo y sus implicaciones prácticas.

En estudios de IO importantes se pueden emplear varios meses (o más) para desarrollar, probar e instalar este sistema de computadora. Parte de este esfuerzo incluye el desarrollo e implantación de un proceso de mantenimiento durante su uso futuro. Si con el tiempo cambian las condiciones, este proceso debe incorporar las modificaciones correspondientes al sistema de computadora y al modelo.

Ejemplo. El Recuadro de aplicación de la sección 2.2 describió un estudio de IO que se elaboró para **Continental Airlines** que condujo a la formulación de un enorme modelo matemático para reasignar tripulaciones a los vuelos cuando ocurrían interrupciones en el itinerario. Como el modelo debe aplicarse de inmediato cuando hay una interrupción, se desarrolló un *sistema de soporte a las decisiones* llamado *CrewSolver* para incorporar tanto el modelo como una gigantesca memoria de datos almacenados que representaban las operaciones actuales. El CrewSolver permite que un coordinador de tripulaciones ingrese datos relacionados con la interrupción en el itinerario y después utilice una interfaz gráfica con el usuario para solicitar una solución inmediata y encontrar la mejor manera de reasignar tripulaciones a los vuelos.

2.6 IMPLEMENTACIÓN

Una vez desarrollado el sistema para aplicar el modelo, la última etapa de un estudio de IO es implementarlo según lo haya establecido la administración. Esta etapa es crítica, pues aquí y sólo aquí se cosecharán los beneficios del estudio. Por lo tanto, es importante que el equipo de IO participe para asegurar que las soluciones del modelo se traduzcan con exactitud en un procedimiento operativo, y para corregir defectos en la solución que se presenten en cualquier momento.

El éxito de la implementación depende en gran medida del apoyo que proporcionen tanto la alta administración como la gerencia operativa. Es más probable que el equipo de IO obtenga este apoyo si ha mantenido a la administración bien informada y ha fomentado la guía de ésta durante el estudio. La buena comunicación ayuda a asegurar que el estudio logre lo que pretende y, por lo tanto, merezca ponerse en práctica. También proporciona a la administración el sentimiento de que el estudio es suyo y facilita el apoyo para su implementación.

La etapa de implementación incluye varios pasos. Primero, el equipo de IO explica en forma cuidadosa a la administración operativa el nuevo sistema que debe adoptar y su relación con la realidad operativa. A continuación, estos dos grupos comparten la responsabilidad de desarrollar los procedimientos que se requieren para poner el sistema en operación. Después, la administración operativa se encarga de proporcionar una capacitación detallada al personal que participa, y se inicia el nuevo curso de acción. Si tiene éxito, el nuevo sistema se podrá emplear durante algunos años. Con este objetivo en mente, el equipo de IO supervisa la experiencia inicial con base en la acción que se tomó para identificar cualquier modificación que deba hacerse en el futuro.

Durante el periodo de uso del nuevo sistema, es importante continuar con la retroalimentación acerca de su funcionamiento y si los supuestos todavía se cumplen. Cuando ocurren desviaciones significativas de los supuestos originales, el modelo debe ser revisado para determinar si necesita modificaciones. El análisis posóptimo que se realice (vea la sección 2.3) puede ser una guía útil en este proceso de revisión.

Durante la fase de culminación del estudio es apropiado que el equipo de investigación de operaciones *documente* su metodología con suficiente claridad y detalle para que el trabajo pueda *reproducirse.* La posibilidad de obtener una *réplica* debe ser parte del código ético profesional del investigador de operaciones. Esta condición es en especial importante cuando se estudian políticas gubernamentales controvertidas.

Ejemplo. Este ejemplo ilustra cómo, en ocasiones, una fase de implementación exitosa requiere el involucramiento de miles de empleados antes de poner en marcha los nuevos procedimientos. **Samsung Electronics Corp.** inició un importante estudio de IO en marzo de 1996 para desarrollar nuevas metodologías y programar el calendario de aplicaciones que pudieran incrementar el nivel de eficiencia de todo el proceso de manufactura de semiconductores y reducir el nivel de inventarios de productos en proceso. El estudio continuó por más de cinco años y llegó a su fin en junio de 2001, debido, de manera principal, al gran esfuerzo que requirió la etapa de implementación. El equipo de IO debió obtener el apoyo de muchos integrantes de las áreas de administración, manufactura e ingeniería; dicho apoyo lo obtuvo mediante la capacitación del personal en los principios y la lógica de los nuevos procedimientos de manufactura. Al final del proceso, más de 3 000 personas habían asistido a las sesiones de capacitación. Después, los nuevos procedimientos fueron incorporados de manera gradual para construir confianza en ellos. Este paciente proceso de implementación pagó grandes dividendos. Los nuevos procedimientos hicieron que la compañía se convirtiera de la manufacturera menos eficiente de la industria de los semiconductores en la más eficiente. Este logro incrementó las ganancias en más de 1 000 millones de dólares en el momento en que se completó el estudio de IO. (La referencia seleccionada A11 describe el estudio a detalle.)

2.7 CONCLUSIONES

Aunque el resto de este libro se enfoca de manera fundamental en la *construcción* y *solución* de modelos matemáticos, en este capítulo se intentó hacer hincapié en que esos elementos sólo son una porción del proceso completo necesario para llevar a cabo un estudio de IO representativo. Las otras etapas que se describieron son también muy importantes para el éxito del estudio. Se pide al lector que en los capítulos subsecuentes no pierda de vista el papel que tienen el modelo y el procedimiento de solución dentro del proceso completo. Después, cuando haya adquirido una comprensión más profunda de los modelos matemáticos, se sugiere que planee regresar y revisar este capítulo con el fin de profundizar en esta perspectiva.

La investigación de operaciones está ligada en forma íntima al empleo de computadoras. Hasta hace poco se usaban de manera casi exclusiva computadoras grandes, pero cada vez es mayor el uso de las computadoras personales y estaciones de trabajo para resolver modelos de IO.

Para concluir la presentación de las etapas más importantes de un estudio de IO, debe observarse que existen muchas excepciones a las “reglas” prescritas en este capítulo. Por su naturaleza, la investigación de operaciones requiere una gran dosis de ingenio e innovación, por lo que es imposible prescribir un procedimiento estándar que los equipos de IO deban seguir siempre. En su lugar, la descripción anterior debe verse como un modelo que representa, a grandes rasgos, cómo llevar a cabo un estudio exitoso de investigación de operaciones.

REFERENCIAS SELECCIONADAS

1. Board, J., C. Sutcliffe y W. T. Ziemba: “Applying Operations Research Techniques to Financial Markets”, en *Interfaces*, **33**(2): 12-24, marzo-abril, 2003.
2. Bradley P. S., U. M. Fayyad y O. L. Mangasarian, “Mathematical Programming for Data Mining: Formulations and Challenges”, en *INFORMS Journal of Computing*, **11**(3): 217-238, verano de 1999.
3. Gass, S. I., “Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis”, en *Operations Research,* **31:** 603-631, 1983.
4. Gass, S. I., “Model World: Danger, Beware the User as Modeler”, en *Interfaces*, **20**(3): 60-64, mayo-junio de 1990.
5. Hall, R. W., “What’s So Scientific about MS/OR?”, en *Interfaces*, **15**(2): 40-45, marzo-abril de 1985.
6. Howard, R. A., “The Ethical OR/MS Professional”, en *Interfaces*, **31**(6): 69-82, noviembre-diciembre de 2001.

7. Miser, H. J., “The Easy Chair: Observation and Experimentation”, en *Interfaces*, **19**(5): 23-30, septiembre-octubre de 1989.
8. Morris, W. T., “On the Art of Modeling”, en *Management Science*, **13**: B707-717, 1967.
9. Murphy, F. H., “The Occasional Observer: Some Simple Precepts for Project Success”, en *Interfaces*, **28**(5): 25-28, septiembre-octubre de 1998.
10. Murphy, F. H., “ASP, The Art and Science of Practice: Elements of the Practice of Operations Research: A Framework”, en *Interfaces*, **35**(2): 154-163, marzo-abril de 2005.
11. Pidd, M., “Just Modeling Through: A Rough Guide to Modeling”, en *Interfaces*, **29**(2): 118-132, marzo-abril de 1999.
12. Williams, H. P., *Model Building in Mathematical Programming*, 4a. ed., Wiley, Nueva York, 1999.
13. Wright, P. D., M. J. Liberatore y R. L. Nydick, “A Survey of Operations Research Models and Applications in Homeland Security”, en *Interfaces* **36**(6): 514-529, noviembre-diciembre de 2006.

Algunas aplicaciones ganadoras de premios con el enfoque de modelado de IO

(En nuestro sitio web www. mhhe.com/hillier se proporciona un vínculo a todos estos artículos.)

A1. Alden, J. M., L. D. Burns, T. Costy, R. D. Hutton, C. A. Jackson, D. S. Kim, K. A. Kohls, J. H. Owen, M. A. Turnquist y D. J. V. Veen, “General Motors Increase Its Production Throughput”, en *Interfaces*, **36**(1): 6-25, enero-febrero de 2006.

A2. Altschuler, S., D. Batavia, J. Bennett, R. Labe, B. Liao, R. Nigam y J. Oh, “Pricing Analysis for Merrill Lynch Integrated Choice”, en *Interfaces*, **32**(1): 5-19, enero-febrero de 2002.

A3. Bixby, A., B. Downs y M. Self, “A Scheduling and Capable-to-Promise Application for Swift & Company”, en *Interfaces*, **36**(1): 69-86, enero-febrero de 2006.

A4. Braklow, J. W., W. W. Graham, S. M. Hassler, K. E. Peck y W. B. Powell, “Interactive Optimization Improves Service and Performance for Yellow Freight System”, en *Interfaces*, **22**(1): 147-172, enero-febrero de 1992.

A5. Cohen, M., P. V. Kamesam, P. Kleindorfer, H. Lee y A. Tekerian, “Optimizer: IBM’s Multi-Echelon Inventory System for Managing Service Logistics”, en *Interfaces*, **20**(1): 65-82, enero-febrero de 1990.

A6. DeWitt, C. W., L. S. Lasdon, A. D. Waren, D. A. Brenner y S. A. Meelhem, “OMEGA: An Improved Gasoline Blending System for Texaco”, en *Interfaces*, **19**(1): 85-101, enero-febrero de 1990.

A7. Goeller, B. F. y el equipo de PAWN, “Planning the Netherlands’ Water Resources”, en *Interfaces*, enero-febrero de 1985.

A8. Hicks, R., R. Madrid, C. Milligan, R. Pruneau, M. Kanaley, Y. Dumas, B. Lacroix, J. Desrosiers y F. Soumis, “Bombardier Flexjet Significantly Improves Its Fracctional Aircraft Ownership Operations”, en *Interfaces*, **35**(1): 49-60, enero-febrero de 2005.

A9. Kaplan, E. H. y E. O’Keefe, “Let the Needles Do the Talking! Evaluating the New Haven Needle Exchange”, en *Interfaces*, **23**(1): 7-26, enero-febrero de 1993.

A10. Kok, T. de, F. Janssen, J. van Doremalen, E. van Wachem, M. Clerkx y W. Peeters, “Philips Electronics Synchronizes Its Supply Chain to End the Bullwhip Effect”, en *Interfaces*, **35**(1): 37-48, enero-febrero de 2005.

A11. Leachman, R. C., J. Kang y V. Lin, “SLIM: Short Cycle Time and Low Inventory in Manufacturing at Samsung Electronics”, en *Interfaces*, **32**(1): 61-77, enero-febrero de 2002.

A12. Taylor, P. E. y S. J. Huxley, “A Break from Tradition for the San Francisco Police: Patrol Officer Scheduling Using an Optimization-Based Decision Support System”, en Interfaces, **19**(1): 4-24, enero-febrero de 1989.

PROBLEMAS

2.1-1. El ejemplo de la sección 2-1 resume un estudio de IO ganador de premios otorgados a Merrill Lynch. Lea la referencia seleccionada A2 que describe este estudio a detalle.

a) Resuma los antecedentes que llevaron a emprender este estudio.

b) Cite el enunciado de una sola oración con la misión general del grupo de IO (llamado el grupo de ciencia de la administración) que realizó este estudio.

c) Explique el tipo de datos que el grupo de ciencia de la administración obtuvo de cada cliente.

d) Identifique las opciones de nuevos precios que fueron proporcionadas a los clientes de la compañía como resultado de este estudio.

e) ¿Cuál es el efecto resultante sobre la posición competitiva de Merrill Lynch?

2.1-2. Lea la referencia seleccionada A1 que describe un estudio de IO ganador de premios realizado para General Motors.
***a*)** Resuma los antecedentes que llevaron a emprender este estudio.
***b*)** ¿Cuál es la meta de este estudio?
***c*)** Describa cómo se utilizó el software para automatizar la recolección de los datos necesarios.
***d*)** ¿Cuánto generó en ahorros documentados y ganancias adicionales la mejora en la producción que resultó de este estudio?

2.1-3. Lea la referencia seleccionada A12 que describe un estudio de IO realizado para el Departamento de Policía de San Francisco.
***a*)** Resuma los antecedentes que llevaron a emprender este estudio.
***b*)** Defina parte del problema a que se refieren las seis metas del sistema de programación a desarrollar.
***c*)** Describa cómo fueron recolectados los datos necesarios.
***d*)** Enumere los distintos beneficios tangibles e intangibles resultado del estudio.

2.1-4. Lea la referencia seleccionada A9 que describe un estudio de IO que se realizó para el Departamento de Salud de New Haven, Connecticut.
***a*)** Resuma los antecedentes que llevaron a emprender este estudio.
***b*)** Describa el sistema que se desarrolló para rastrear y probar cada aguja y cada jeringa con el fin de reunir los datos necesarios.
***c*)** Resuma los resultados iniciales de este sistema de rastreo y pruebas.
***d*)** Describa el efecto real y potencial de este estudio en las políticas gubernamentales.

2.2-1. Lea el artículo al que se hace referencia en el Recuadro de aplicación de la sección 2.2, el cual describe por completo el estudio de IO que se resume en dicho recuadro. Enumere los diferentes beneficios financieros y no financieros que resultaron de este estudio.

2.2-2. Lea la referencia seleccionada A3 que describe un estudio de IO realizado para Swift & Company.
***a*)** Resuma los antecedentes que llevaron a emprender este estudio.
***b*)** Describa el propósito de cada uno de los tres tipos generales de modelos que se formularon durante este estudio.
***c*)** ¿Cuántos modelos específicos utiliza en la actualidad la compañía como resultado de este estudio?
***d*)** Enumere los distintos beneficios financieros y no financieros que resultaron de este estudio.

2.2-3. Lea la referencia seleccionada A7 que describe un estudio de IO que se realizó para el Rijkswaterstaat de Holanda. (Preste especial atención en las páginas 3-20 y 30-32.)
***a*)** Resuma los antecedentes que llevaron a emprender este estudio.
***b*)** Resuma el objetivo de cada uno de los cinco modelos matemáticos que se describen en las páginas 10-18.
c) Resuma las "medidas de efecto" (medidas de desempeño) para comparar las políticas que se describen en las páginas 6-7 de este artículo.
***d*)** Enumere los distintos beneficios tangibles e intangibles que resultaron de este estudio.

2.2-4. Lea la referencia seleccionada 5.
***a*)** Identifique el ejemplo del autor sobre un modelo de las ciencias naturales y uno de IO.
***b*)** Describa el punto de vista del autor sobre la manera en que los principios básicos del uso de modelos para realizar investigación en ciencias naturales puede usarse para guiar la *investigación sobre las operaciones* (IO).

2.3-1. Lea la referencia seleccionada A10 que describe un estudio de IO realizado para Philips Electronics.
***a*)** Resuma los antecedentes que llevaron a realizar este estudio.
***b*)** ¿Cuál fue el propósito de este estudio?
***c*)** ¿Cuáles fueron los beneficios de desarrollar software para apoyar la resolución más rápida de problemas?
***d*)** Enumere los cuatro pasos del proceso de colaboración y planeación que resultó de este estudio.
***e*)** Enumere los distintos beneficios financieros y no financieros que resultaron de este estudio.

2.3-2. Lea la referencia seleccionada 5.
***a*)** Describa el punto de vista del autor sobre el hecho de que la única meta al usar un modelo debe ser encontrar su solución óptima.
***b*)** Resuma el punto de vista del autor sobre los papeles complementarios del modelado, la evaluación de la información que se obtiene y la aplicación del juicio del tomador de decisiones para determinar un curso de acción.

2.4-1. Consulte las páginas 18-20 de la referencia seleccionada A7 que describe un estudio de IO realizado para el Rijkswaterstaat de Holanda. Describa una lección importante aprendida con la validación del modelo en este estudio.

2.4-2. Lea la referencia seleccionada 7. Resuma el punto de vista del autor sobre el papel de la observación y la experimentación en el proceso de validación del modelo.

2.4-3. Lea las páginas 603-617 de la referencia seleccionada 3.
***a*)** ¿Qué dice el autor sobre el hecho de que un modelo se puede validar por completo?
***b*)** Resuma la diferencia entre *validación del modelo*, *de los datos*, *validación lógica/matemática*, *predictiva*, *operativa* y *dinámica*.
***c*)** Describa el papel del *análisis de sensibilidad* en la *validación operativa* de un modelo.
***d*)** ¿Qué dice el autor sobre la existencia de una metodología de validación adecuada para todos los modelos?
***e*)** Cite la página del artículo que enumera los pasos básicos de la validación.

2.5-1. Lea la referencia seleccionada A6 que describe un estudio de IO que se realizó para Texaco.
***a*)** Resuma los antecedentes que llevaron a emprender este estudio.
***b*)** Describa brevemente la interfaz del usuario con el sistema de apoyo a las decisiones OMEGA que se desarrolló como resultado de este estudio.
***c*)** OMEGA se actualiza y amplía en forma constante para reflejar los cambios en el ambiente de las operaciones. Describa los distintos tipos de cambios realizados.
***d*)** Resuma cómo se usa el sistema OMEGA.
***e*)** Enumere los distintos beneficios tangibles e intangibles que resultaron de este estudio.

2.5-2. Consulte la referencia seleccionada A4 que describe un estudio de IO que se realizó para la Yellow Freight System, Inc.
***a*)** Con base en lo expuesto en las páginas 147-149 de este artículo, resuma los antecedentes que llevaron a emprender este estudio.
***b*)** De acuerdo con lo que se explicó en la página 150, describa en forma breve el sistema de computadora SYSNET que se desarro-

lló como resultado de este estudio. También dé un resumen de las aplicaciones de SYSNET.

c) Según lo expuesto en las páginas 162-163, describa por qué eran importantes los aspectos *interactivos* de SYSNET.

d) Con base en lo que se explicó en la página 163, resuma la salida de SYSNET.

e) De acuerdo con lo que se expone en las páginas 168-172, resuma los distintos beneficios que ha generado el uso de SYSNET.

2.6-1. Consulte las páginas 163-167 de la referencia seleccionada A4 que describe el estudio de IO que se realizó para la Yellow Freight System, Inc. y el sistema de computadora SYSNET que se desarrolló como consecuencia de él.

a) Describa de modo breve la forma en que el equipo de IO logró el apoyo de la alta administración para la implantación de SYSNET.

b) Describa de forma breve la estrategia de implementación desarrollada.

c) Describa de manera somera la implementación en el campo de trabajo.

d) Describa de modo sucinto cómo se usaron los incentivos y la vigilancia de la administración para implementar SYSNET.

2.6-2. Lea la referencia seleccionada A5 que describe un estudio de IO que se realizó para IBM y el sistema Optimizer que resultó de él.

a) Resuma los antecedentes que llevaron a emprender este estudio.

b) Enumere los factores de complicación a los que se enfrentaron los integrantes del equipo de IO al inicio del desarrollo de un modelo y un algoritmo de solución.

c) Presente una descripción breve de las pruebas anteriores a la implementación de Optimizer.

d) Describa en forma breve las pruebas de implementación en el campo de trabajo.

e) Describa de manera sucinta el proceso de implementación a nivel nacional.

f) Enumere los distintos beneficios tangibles e intangibles que resultaron de este estudio.

2.7-1. De la última parte de las referencias seleccionadas que se presentan al final del capítulo, seleccione una de las aplicaciones ganadoras de premios del enfoque de modelado de IO (sin considerar las asignadas para otros problemas). Lea el artículo y escriba un resumen de dos páginas de la aplicación y los beneficios que proporcionó (incluya los beneficios no financieros).

2.7-2. De la última parte de las referencias seleccionadas que se presentan al final del capítulo, seleccione tres de las aplicaciones ganadoras de premios del enfoque de modelado de IO (sin considerar las asignadas para otros problemas). Lea cada uno de los artículos y escriba un resumen de una página de la aplicación y los beneficios que proporcionó (incluya los beneficios no financieros).

2.7-3. Lea la referencia seleccionada 4. El autor describe 13 etapas detalladas de cualquier estudio que desarrolla y aplica un modelo basado en computadora, mientras que este capítulo describe seis etapas más amplias. Establezca una relación entre los dos niveles al enumerar las etapas detalladas que caen, parcial o primordialmente, dentro de cada etapa más amplia.

CAPÍTULO 3

Introducción a la programación lineal

El desarrollo de la programación lineal ha sido clasificado como uno de los avances científicos más importantes de mediados del siglo XX, y estamos de acuerdo con esta aseveración. Su efecto desde 1950 ha sido extraordinario. En la actualidad es una herramienta de uso normal que ha ahorrado miles o millones de dólares a muchas compañías o negocios, incluso empresas medianas, en los distintos países industrializados del mundo; su aplicación a otros sectores de la sociedad se ha ampliado con rapidez. Una proporción muy grande de los programas científicos en computadoras está dedicada al uso de la programación lineal. Se han escrito docenas de libros de texto sobre esta materia y se cuentan por cientos los artículos *publicados* que describen aplicaciones importantes.

¿Cuál es la naturaleza de esta notable herramienta y qué tipos de problemas puede manejar? El lector adquirirá una noción de este tema a medida que trabaje en los ejemplos que se presentarán más adelante. Sin embargo, un resumen verbal puede permitirle elaborar una idea. Expresado en forma breve, el tipo más común de aplicación abarca el problema general de asignar de la mejor manera posible —es decir, de *forma óptima*— *recursos limitados* a *actividades que compiten* entre sí por ellos. Con más precisión, este problema consiste en elegir el nivel de ciertas actividades que compiten por recursos escasos necesarios para realizarlas. Después, los niveles de actividad que se eligen dictan la cantidad de recursos que consumirá cada una de ellas. La variedad de situaciones a las que se puede aplicar esta descripción es sin duda muy grande, ya que abarca desde la asignación de instalaciones de producción a los productos hasta la asignación de los recursos nacionales a las necesidades de un país; desde la selección de una cartera de inversiones hasta la selección de los patrones de envío; desde la planeación agrícola hasta el diseño de una terapia de radiación, etc. No obstante, el ingrediente común de todas estas situaciones es la necesidad de asignar recursos a las actividades mediante la elección de los niveles de éstas.

La programación lineal utiliza un modelo matemático para describir el problema. El adjetivo *lineal* significa que todas las funciones matemáticas del modelo deben ser *funciones lineales*. En este caso, la palabra *programación* no se refiere aquí a términos computacionales; en esencia es sinónimo de *planeación*. Por lo tanto, la programación lineal involucra la *planeación de actividades* para obtener un resultado óptimo; esto es, el resultado que mejor alcance la meta especificada —de acuerdo con el modelo matemático— entre todas las alternativas factibles.

Aunque la asignación de recursos a las actividades es la aplicación más frecuente, la programación lineal tiene muchas otras posibilidades. En realidad, *cualquier* problema cuyo modelo matemático se ajuste al formato general del modelo de programación lineal, es un problema de programación lineal. (Por esta razón, un problema de programación lineal y su modelo se denominan con frecuencia *programa lineal*, o incluso sólo *PL*.) Aún más, se dispone de un procedimiento de solución muy eficiente llamado **método símplex** para resolver estos problemas lineales, incluso los de gran tamaño. Éstas son algunas razones del tremendo efecto de la programación lineal en las décadas recientes.

Debido a su gran importancia hemos dedicado a la programación lineal éste y los siguientes seis capítulos. Después de presentar aquí las características generales de programación lineal, los capítulos 4 y 5 se dedican al método símplex. El capítulo 6 analiza los problemas de programación

lineal *después* de la aplicación inicial del método símplex. El capítulo 7 examina varias extensiones muy empleadas de este método e introduce el *algoritmo de punto interior* que en ocasiones se usa para resolver problemas de programación lineal aún más grandes que los que maneja el método símplex. Los capítulos 8 y 9 consideran algunos problemas especiales de programación lineal cuya trascendencia justifica su estudio individual.

Además, en varios de los capítulos posteriores se verán aplicaciones de programación lineal a otras áreas de la investigación de operaciones.

Este capítulo comienza con el desarrollo de un ejemplo prototípico simplificado de un problema de programación lineal. Este ejemplo es tan pequeño que puede resolverse de manera directa en una gráfica. En las secciones 3.2 y 3.3 se presentan el *modelo general de programación lineal* y sus supuestos básicos. La sección 3.4 proporciona algunos ejemplos adicionales de programación lineal. La sección 3.5 describe cómo pueden establecerse y resolverse problemas de programación lineal de tamaño mediano en una hoja de cálculo. Sin embargo, algunos problemas reales requieren modelos en verdad *masivos*. La sección 3.6 ilustra cómo suelen surgir estos modelos de gran tamaño y cómo se pueden formular de manera correcta con la ayuda de lenguajes especiales de modelado como MPL —su formulación se describe en esta sección— o LINGO (la formulación de este modelo se presenta en el suplemento 2 de este capítulo en el sitio web del libro).

3.1 EJEMPLO PROTOTÍPICO

La WYNDOR GLASS CO. produce artículos de vidrio de alta calidad, entre ellos ventanas y puertas de vidrio. Tiene tres plantas. Los marcos y molduras de aluminio se hacen en la planta 1, los de madera en la planta 2; la 3 produce el vidrio y ensambla los productos.

Debido a una reducción de las ganancias, la alta administración ha decidido reorganizar la línea de producción de la compañía. Se discontinuarán varios productos no rentables y se dejará libre una parte de la capacidad de producción para emprender la fabricación de dos productos nuevos cuyas ventas potenciales son muy prometedoras:

Producto 1: una puerta de vidrio de 8 pies con marco de aluminio
Producto 2: una ventana corrediza con marco de madera de 4 por 6 pies

El producto 1 requiere parte de la capacidad de producción en las plantas 1 y 3 y nada en la planta 2. El producto 2 sólo necesita trabajo en las plantas 2 y 3. La división de comercialización ha concluido que la compañía puede vender todos los productos que se puedan fabricar en las plantas. Sin embargo, como ambos productos competirían por la misma capacidad de producción en la planta 3, no está claro cuál *mezcla* de productos sería la *más rentable*. Por lo tanto, se ha formado un equipo de IO para estudiar este problema.

El grupo comenzó por realizar juntas con la alta administración para identificar los objetivos del estudio. Como consecuencia de ellas se desarrolló la siguiente definición del problema:

> Determinar cuáles *tasas de producción* deben tener los dos productos con el fin de *maximizar las utilidades totales*, sujetas a las restricciones impuestas por las capacidades de producción limitadas disponibles en las tres plantas. (Cada producto se fabricará en lotes de 20 unidades, de manera que la *tasa de producción* está definida como el número de lotes que se producen a la semana.) Se permite *cualquier* combinación de tasas de producción que satisfaga estas restricciones, incluso no fabricar uno de los productos y elaborar todo lo que sea posible del otro.

El equipo de IO también identificó los datos que necesitaba reunir:

1. Número de horas de producción disponibles por semana en cada planta para fabricar estos nuevos productos. (Casi todo el tiempo de estas plantas está comprometido con los productos actuales, lo que limita la capacidad para manufacturar nuevos productos.)
2. Número de horas de fabricación que se emplea para producir cada lote de cada artículo nuevo en cada una de las plantas.
3. La ganancia por lote de cada producto nuevo. (Se escogió la *ganancia por lote producido* como una medida adecuada una vez que el equipo llegó a la conclusión de que la ganancia incremental de cada lote adicional producido sería, en esencia, *constante*, sin que importase el número total de lotes producidos. Debido a que no se incurre en costos sustanciales para iniciar

Recuadro de aplicación

Swift & Company es una empresa diversificada productora de proteína con base en Greeley, Colorado. Con ventas anuales de más de 8 000 millones de dólares, la carne de res y sus productos derivados son, por mucho, la parte más grande del negocio de la compañía.

A fin de mejorar las ventas de la empresa y su desempeño en la manufactura, la alta administración concluyó que necesitaba alcanzar tres objetivos importantes. Uno fue permitir a los representantes de servicio al cliente hablar a sus más de 8 000 clientes para transmitirles información precisa acerca de la disponibilidad de inventario actual y futuro, al mismo tiempo que consideraban fechas de entrega solicitadas y edad máxima del producto en el momento de su entrega. Un segundo objetivo fue producir un programa eficiente de nivel de turno para cada planta en un horizonte de 28 días. El tercer objetivo consistió en determinar de manera exacta si una planta podía embarcar una cantidad solicitada de pedidos-líneas-artículos en la fecha y a la hora requeridas dadas la disponibilidad de ganado y las restricciones impuestas por la capacidad de la planta.

Para enfrentar estos tres desafíos, un equipo de IO desarrolló un *sistema integrado de 45 modelos de programación lineal* basado en tres formulaciones de modelo para programar de manera dinámica sus operaciones de fabricación de carne en cinco plantas en tiempo real cuando recibe los pedidos. *Los beneficios totales auditados que se observaron en el primer año* de operación de este sistema fueron de **12.74 millones de dólares**, de los cuales 12 millones correspondieron a la *optimización de la mezcla de productos*. Entre otros beneficios se destacan la disminución de las órdenes perdidas, la reducción de los descuentos de precio y la mejora de las entregas a tiempo.

Fuente: A. Bixby, B. Downs y M. Self, "A Scheduling and Capable-to-Promise Application for Swift & Company", en *Interfaces*, **36**(1): 39-50, enero-febrero de 2006. (En nuestro sitio web, www.mhhe.com/hillier, se proporciona un vínculo con este artículo.)

la producción y la comercialización de estos nuevos productos, la ganancia total de cada uno es aproximadamente la *ganancia por lote que se produce* multiplicada por el *número de lotes*.)

La obtención de estimaciones razonables de estas cantidades requirió del apoyo de personal clave en varias unidades de la compañía. El personal de la división de manufactura proporcionó los datos de la primera categoría mencionada. En la segunda categoría, el desarrollo de estimaciones requirió un análisis de los ingenieros de manufactura involucrados en el diseño de los procesos de producción para elaborar los nuevos artículos. Al analizar los datos de costos que se obtuvieron, junto con la decisión sobre los precios de la división de marketing, el departamento de contabilidad calculó las estimaciones para la tercera categoría.

La tabla 3.1 resume los datos reunidos.

De inmediato, el equipo de IO reconoció que se trataba de un problema de programación lineal del tipo clásico de **mezcla de productos** y procedió a la formulación del modelo matemático correspondiente.

Formulación como un problema de programación lineal

La definición del problema planteado indica que las decisiones que deben tomarse son el número de lotes de los productos que se fabricarán semanalmente, de manera que se maximice su ganancia total.

Para formular el modelo matemático de programación lineal de este problema se define

x_1 = número de lotes del producto 1 que se fabrican por semana
x_2 = número de lotes del producto 2 que se fabrican por semana
Z = ganancia semanal total (en miles de dólares) que generan estos dos productos

TABLA 3.1 Datos del problema de la Wyndor Glass Co.

	Tiempo de producción por lote, horas		
	Producto		
Planta	1	2	Tiempo de producción disponible a la semana, horas
1	1	0	4
2	0	2	12
3	3	2	18
Ganancia por lote	$3 000	$5 000	

Por lo tanto, x_1 y x_2 son las *variables de decisión* del modelo. Si se usa el último renglón de la tabla 3.1 se obtiene

$$Z = 3x_1 + 5x_2.$$

El objetivo es elegir los valores de x_1 y x_2 que *maximice* $Z = 3x_1 + 5x_2$, sujeta a las restricciones impuestas sobre sus valores por las capacidades de producción limitadas de las cuales se disponen en las tres plantas. La tabla 3.1 indica que cada lote del producto 1 que se produce por semana emplea una hora de producción en la planta 1, y sólo se dispone de 4 horas semanales. En términos matemáticos, esta restricción se expresa mediante la desigualdad $x_1 \leq 4$. De igual manera, la planta 2 impone la restricción $2x_2 \leq 12$. El número de horas de producción usadas a la semana en la planta 3 que se consume al elegir x_1 y x_2 como las tasas de producción de los nuevos productos sería $3x_1 + 2x_2$. En consecuencia, la expresión matemática de la restricción de la planta 3 es $3x_1 + 2x_2 \leq 18$. Por último, como las tasas de producción no pueden ser negativas, es necesario restringir las variables de decisión a valores no negativos: $x_1 \geq 0$ y $x_2 \geq 0$.

Para resumir, en el lenguaje matemático de programación lineal, el problema consiste en seleccionar valores de x_1 y x_2 para

$$\text{Maximizar} \qquad Z = 3x_1 + 5x_2,$$

sujeta a las restricciones

$$\begin{aligned} x_1 &\leq 4 \\ 2x_2 &\leq 12 \\ 3x_1 + 2x_2 &\leq 18 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

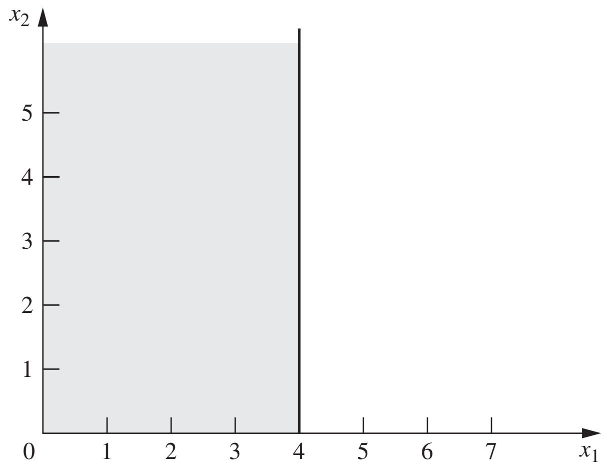
(Observe cómo la información de la tabla 3.1 en esencia se duplica en la distribución de los coeficientes de x_1 y x_2 en el modelo de programación lineal.)

Solución gráfica

Este pequeño problema tiene sólo dos variables de decisión, esto es, sólo dos dimensiones, así que se puede usar un procedimiento gráfico para resolverlo. Este procedimiento incluye la construcción de una gráfica de dos dimensiones con x_1 y x_2 como los ejes. El primer paso es identificar los valores de (x_1, x_2) permitidos por las restricciones. Este objetivo se logra dibujando cada una de las rectas que limitan los valores permitidos por una restricción. Para comenzar, observe que las restricciones de no negatividad $x_1 \geq 0$ y $x_2 \geq 0$ exigen que el punto (x_1, x_2) se encuentre en el lado *positivo* de los ejes (incluso sobre *cualquiera* de los dos ejes), es decir, en el primer cuadrante. Después, debe observarse que la restricción $x_1 \leq 4$ significa que (x_1, x_2) no puede estar a la derecha de la recta $x_1 = 4$. Estos resultados se muestran en la figura 3.1, en la que el área sombreada contiene los únicos valores de (x_1, x_2) permitidos.

De manera parecida, la restricción $2x_2 \leq 12$ (o de modo equivalente, $x_2 \leq 6$) implica que la recta $2x_2 = 12$ debe agregarse a la frontera de la región permisible. La última restricción, $3x_1 + 2x_2 \leq 18$, se encuentra al graficar los puntos (x_1, x_2) tales que $3x_1 + 2x_2 = 18$ (otra recta) para completar la frontera. (Observe que los puntos que cumplen $3x_1 + 2x_2 \leq 18$ son aquellos que están sobre o por debajo de la recta $3x_1 + 2x_2 = 18$, por lo que ésta es la recta que limita, y más allá de ella, la desigualdad no se satisface.) En la figura 3.2 se muestra la región de valores permisibles de (x_1, x_2), llamada **región factible**. (La demostración llamada *Graphical Method* —método gráfico— en el OR Tutor proporciona un ejemplo detallado de la construcción de la región factible.)

El paso final es seleccionar, dentro de esta región factible, el punto que maximiza el valor de $Z = 3x_1 + 5x_2$. Para descubrir cómo realizar este paso de manera eficiente se pueden intentar algunos valores por prueba y error. Por ejemplo, probar, $Z = 10 = 3x_1 + 5x_2$ para ver si existe algún valor de (x_1, x_2) dentro de la región permisible que dé un valor de 10 para Z. Si se dibuja la recta $3x_1 + 5x_2 = 10$ se puede ver que existen muchos puntos sobre esta recta que están dentro de la región (vea la figura 3.3). Después de intentar este valor arbitrario de $Z = 10$ se tiene una mejor



■ **FIGURA 3.1**
El área sombreada muestra los valores de (x_1, x_2) permitidos por $x_1 \geq 0$, $x_2 \geq 0$, $x_1 \leq 4$.

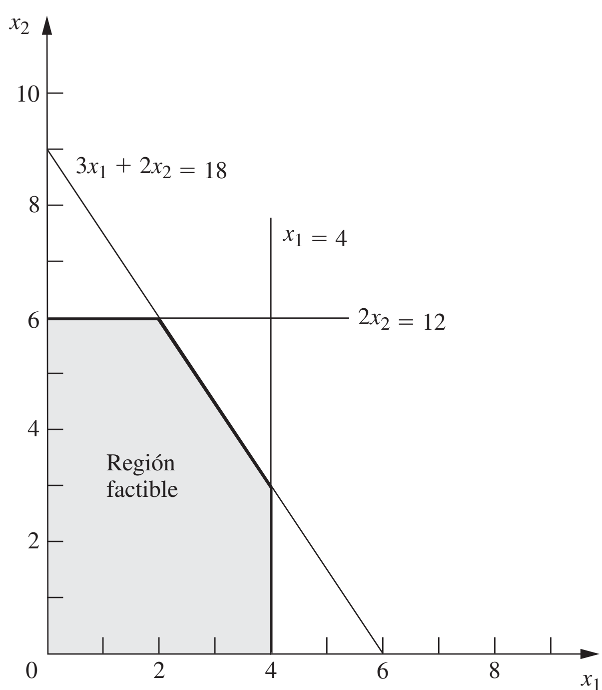
perspectiva, debe intentarse ahora un valor arbitrario más grande, por ejemplo, $Z = 20 = 3x_1 + 5x_2$. De nuevo, la figura 3.3 revela que un segmento de la recta $3x_1 + 5x_2 = 20$ se encuentra dentro de la región, de manera que el máximo valor permisible de Z debe ser, por lo menos, 20.

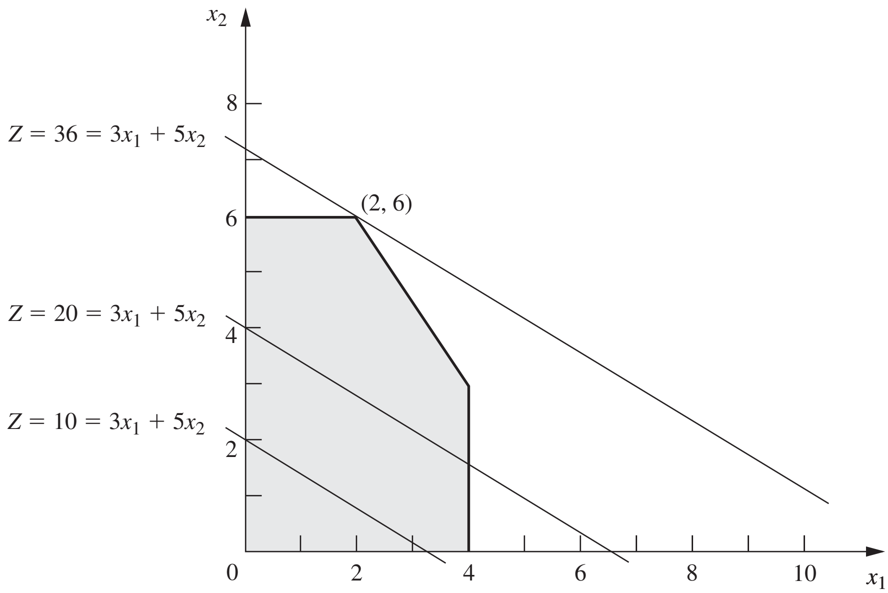
Observe ahora en la figura 3.3 que las dos rectas que se acaban de graficar son paralelas. Esto no es coincidencia, ya que *cualquier* recta construida de esta manera tiene la forma $Z = 3x_1 + 5x_2$ para el valor seleccionado de Z, lo que implica que $5x_2 = -3x_1 + Z$ o, en forma equivalente,

$$x_2 = -\frac{3}{5}x_1 + \frac{1}{5}Z$$

Esta última ecuación, llamada forma de **pendiente-ordenada al origen** de la función objetivo, demuestra que la *pendiente* de las rectas es $-\frac{3}{5}$ (ya que cada incremento de una unidad en x_1 hace que x_2 cambie en $-\frac{3}{5}$), mientras que la ordenada al origen de la recta —la *intersección* con el eje x_2—

■ **FIGURA 3.2**
El área sombreada muestra los valores permitidos de (x_1, x_2), llamada la región factible.





■ **FIGURA 3.3**
El valor de (x_1, x_2) que maximiza $3x_1 + 5x_2$ es (2, 6).

es $\frac{1}{5}Z$ (puesto que $x_2 = \frac{1}{5}Z$ cuando $x_1 = 0$). El hecho de que la pendiente esté fija en $-\frac{3}{5}$ significa que *todas* las rectas construidas de esta manera son paralelas.

De nuevo, si se comparan las rectas $10 = 3x_1 + 5x_2$ y $20 = 3x_1 + 5x_2$ en la figura 3.3, es posible observar que la recta que da el valor mayor de Z ($Z = 20$) se encuentra más lejos del origen hacia arriba que la otra recta ($Z = 10$). Este hecho también está implícito en la forma de pendiente-ordenada al origen de la función objetivo, lo que indica que la intersección con el eje x_1 ($\frac{1}{5}Z$) aumenta cuando crece el valor seleccionado de Z.

Estas observaciones implican que el procedimiento de prueba y error para construir las rectas de la figura 3.3 involucra sólo dibujar una familia de rectas paralelas que contengan al menos un punto en la región factible y elegir la que corresponda al mayor valor de Z. La figura 3.3 muestra que esta recta pasa por el punto (2, 6), lo cual indica que la **solución óptima** es $x_1 = 2$ y $x_2 = 6$. La ecuación de esta recta es $3x_1 + 5x_2 = 3(2) + 5(6) = 36 = Z$, lo cual indica que el valor óptimo de Z es $Z = 36$. El punto (2, 6) se encuentra en la intersección de las dos rectas $2x_2 = 12$ y $3x_1 + 2x_2 = 18$, que se muestra en la figura 3.2, por lo que el punto se puede calcular de manera algebraica como la solución simultánea de estas dos ecuaciones.

Una vez estudiado el procedimiento de prueba y error para encontrar el punto óptimo (2, 6) es posible seguir los pasos de este método en otros problemas. En lugar de dibujar varias rectas paralelas, es suficiente marcar una de ellas con una regla para establecer la pendiente y después mover la regla con pendiente fija sobre la región factible en la dirección en que Z mejora. (Cuando el objetivo sea *minimizar Z* la regla deberá moverse en la dirección en que *Z decrece*.) La regla se deja de mover en el momento en que todavía pasa por un punto de esta región. Este punto es la *solución óptima* deseada.

Con frecuencia se hace referencia a este procedimiento como el **método gráfico** de programación lineal. Se puede usar para resolver cualquier problema de programación lineal con dos variables de decisión. Con alguna dificultad es posible extender el método a tres variables de decisión, pero no más de tres. (En el siguiente capítulo se estudia el *método símplex* para resolver problemas más grandes.)

Conclusiones

El equipo de IO utilizó este procedimiento para encontrar que la solución óptima deseada es $x_1 = 2$, $x_2 = 6$, con $Z = 36$. Esta solución indica que la Wyndor Glass Co. debe fabricar los productos 1 y 2 a una tasa de 2 y 6 lotes por semana, respectivamente, con una ganancia total resultante de 36 000 dólares semanales. No existe otra mezcla de los dos productos que sea tan redituable, *de acuerdo con el modelo*.

No obstante, en el capítulo 2 se puso de manifiesto que un buen estudio de investigación de operaciones no sólo encuentra *una* solución para el modelo *inicial* formulado. Cada una de las seis etapas que se describieron es importante, incluso las pruebas exhaustivas del modelo (vea la sección 2.4) y el análisis posóptimo (sección 2.3).

Si reconoce la totalidad de estas realidades prácticas, el equipo de IO está listo para evaluar la validez del modelo de una manera más crítica (esta explicación continuará en la sección 3.3), y para llevar a cabo un análisis de sensibilidad sobre el efecto que tendría el hecho de que las estimaciones dadas en la tabla 3.1 fueran diferentes debido a inexactitudes, cambios en las circunstancias, etc. (Este tema continuará en la sección 6.7.)

Continuación del proceso de aprendizaje con OR Courseware

Éste es el primero de muchos puntos en los cuales será útil emplear el *OR Courseware* que se encuentra en el CD que acompaña al libro. Un programa clave en este CD es el llamado OR Tutor que contiene un ejemplo de demostración completo del *método gráfico* que se estudia en esta sección. Esta demostración comienza por la introducción de un problema y la formulación de un modelo de programación lineal, antes de aplicar el método gráfico para resolverlo, con la intención de proporcionar un **ejemplo adicional** de formulación de un modelo. Al igual que muchos otros ejemplos de demostración en otras secciones, éste resalta los conceptos que son difíciles de explicar en una página impresa. En el apéndice 1 se puede consultar la documentación sobre el software.

Si el lector desea ver **más ejemplos** puede consultar la sección de problemas resueltos —**Worked Examples**— en el sitio web del libro. Esta sección incluye unos cuantos ejemplos con soluciones completas para casi todos los capítulos, así como un complemento de los ejemplos del libro y del OR Tutor. Los ejemplos de este capítulo comienzan con un problema relativamente directo que implica la formulación de un pequeño modelo de programación lineal y la aplicación del método gráfico. Los ejemplos siguientes implicarán de manera progresiva un reto mayor.

Otra parte clave del OR Courseware es un programa llamado **IOR Tutorial**. Éste realiza muchos procedimientos interactivos para ejecutar los diferentes métodos de solución que se presentan en el libro, lo que permite que el lector se enfoque en el aprendizaje y la ejecución de la lógica del método en forma eficiente, mientras que la computadora realiza los cálculos numéricos. Se incluye un procedimiento interactivo para aplicar el método gráfico en la programación lineal. Una vez que se haya captado este primer procedimiento, un segundo enfoque permite aplicar con rapidez el método gráfico para desarrollar análisis de sensibilidad sobre el efecto de cambios en los datos del problema. Después, es posible imprimir los trabajos y resultados como una tarea. Como los otros procedimientos del IOR Tutorial, éstos están específicamente diseñados para proporcionar al lector una experiencia de aprendizaje eficiente, amena y enriquecedora mientras realiza sus tareas.

Cuando se formule un modelo de programación lineal con más de dos variables de decisión, por lo que no puede usarse el método gráfico, el *método símplex* descrito en el capítulo 4 permitirá encontrar una solución óptima de inmediato. Obtenerla también es útil para la *validación del modelo* puesto que encontrar una *solución sin sentido* indica que se cometieron errores en la formulación del modelo.

En la sección 1.4 se mencionó que el OR Courseware es una introducción a los tres paquetes de software comerciales que más se usan —Excel Solver, LINGO/LINDO y MPL/CPLEX— para resolver una variedad de modelos de IO. Los tres paquetes incluyen el método símplex para resolver problemas de programación lineal. En la sección 3.5 se describe cómo usar Excel para formular y resolver modelos de programación lineal en el formato de una hoja de cálculo. Las descripciones de los otros paquetes se proporcionan en la sección 3.6 —MPL y LINGO—, los suplementos 1 y 2 de este capítulo en el sitio web del libro —LINGO—, la sección 4.8 —CPLEX y LINDO— y el apéndice 4.1 —LINDO—. Además, el OR Courseware incluye un archivo para cada uno de los tres paquetes que muestra cómo se puede usar para resolver los ejemplos en este capítulo.

3.2 MODELO DE PROGRAMACIÓN LINEAL

El problema de la Wyndor Glass Co. se diseñó para ilustrar un problema común de programación lineal, en versión simplificada. Sin embargo, esta técnica es muy versátil como para describirla mediante un solo ejemplo. En esta sección se presentarán las características generales de los problemas de programación lineal y las distintas formas legítimas del modelo matemático.

TABLA 3.2 Terminología común de programación lineal

Ejemplo modelo	Problema general
Capacidad de producción de las plantas 3 plantas	Recursos *m* recursos
Fabricación de productos 2 productos Tasa de producción del producto j, x_j	Actividades *n* actividades Nivel de actividad j, x_j
Ganancia Z	Medida global de desempeño Z

Comenzaremos por establecer la terminología y notación básicas. La primera columna de la tabla 3.2 resume los componentes del problema de la Wyndor Glass Co. La segunda introduce términos más generales de estos componentes, que se ajustarán a muchos problemas de programación lineal. Los términos clave son *recursos* y *actividades* en los que m denota el número de tipos de recursos que se pueden usar y n el número de actividades que se consideran. Algunos ejemplos de recursos son dinero y tipos especiales de maquinaria, equipo, vehículos y personal. Los ejemplos de actividades incluyen inversión en proyectos específicos, publicidad en un medio determinado y el envío de bienes de cierta fuente a cierto destino. En cualquier aplicación de programación lineal es posible que todas las actividades sean de un tipo general (como cualquiera de estos tres ejemplos), a consecuencia de lo cual cada una correspondería en forma individual a las alternativas específicas dentro de esta categoría general.

Como se describió en la introducción del capítulo, el tipo más usual de aplicación de programación lineal involucra la asignación de recursos a ciertas actividades. La cantidad disponible de cada recurso es limitada, de forma que debe asignarse con todo cuidado. La determinación de esta asignación implica elegir los niveles de las actividades que lograrán el mejor valor posible de la *medida global de desempeño*.

Ciertos símbolos se usan de manera convencional para denotar los diversos componentes de un modelo de programación lineal. Estos símbolos se enumeran a continuación, junto con su interpretación para el problema general de asignación de recursos a actividades.

Z = valor de la medida global de desempeño.
x_j = nivel de la actividad j (para $j = 1, 2, \ldots, n$).
c_j = incremento en Z que se obtiene al aumentar una unidad en el nivel de la actividad j.
b_i = cantidad de recurso i disponible para asignarse a las actividades (para $i = 1, 2, \ldots, m$).
a_{ij} = cantidad del recurso i consumido por cada unidad de la actividad j.

El modelo plantea el problema en términos de tomar decisiones sobre los niveles de las actividades, por lo que $x_1, x_2, \ldots, x_n$ se llaman **variables de decisión**. Como se resume en la tabla 3.3, los valores

TABLA 3.3 Datos necesarios para elaborar un modelo de programación lineal para manejar la asignación de recursos a actividades

	Consumo de recursos por unidad de actividad				
	Actividad				
Recurso	**1**	**2**	**. . .**	***n***	**Cantidad de recursos disponibles**
1	a_{11}	a_{12}	. . .	a_{1n}	b_1
2	a_{21}	a_{22}	. . .	a_{2n}	b_2
.					.
.	. . .	. . .	. . .	. . .	.
.					.
m	a_{m1}	a_{m2}	. . .	a_{mn}	b_m
Contribución a Z por unidad de actividad	c_1	c_2	. . .	c_n	

de c_j, b_i y a_{ij} (para $i = 1, 2, \ldots, m$ y $j = 1, 2, \ldots, n$) son las *constantes de entrada* al modelo. Las c_j, b_i y a_{ij} también se conocen como **parámetros** del modelo.

Observe la correspondencia entre la tabla 3.3 y la tabla 3.1.

Una forma estándar del modelo

Para proceder con el problema de la Wyndor Glass Co., ahora se puede formular el modelo matemático del problema general de asignar recursos a actividades. En particular, este modelo consiste en elegir valores de $x_1, x_2, \ldots, x_n$ para

$$\text{Maximizar} \quad Z = c_1x_1 + c_2x_2 + \cdots + c_nx_n,$$

sujeta a las restricciones

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &\le b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &\le b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &\le b_m, \end{aligned}$$

y

$$x_1 \ge 0, \quad x_2 \ge 0, \quad \ldots, \quad x_n \ge 0.$$

Ésta es llamada *nuestra forma estándar*[1] del problema de programación lineal. Cualquier situación cuya formulación matemática se ajuste a este modelo es un problema de programación lineal.

Observe que el modelo del problema de la Wyndor Glass Co. se ajusta a la forma estándar con $m = 3$ y $n = 2$.

En este momento se puede resumir la terminología común de los modelos de programación lineal. La función que se desea maximizar, $c_1x_1 + c_2x_2 + \cdots + c_nx_n$, se llama **función objetivo**. Por lo general, se hace referencia a las limitaciones como **restricciones**. Las primeras m restricciones (aquellas con una *función* de todas las variables $a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n$ en el lado izquierdo) a veces reciben el nombre de **restricciones funcionales** (o *restricciones estructurales*). De manera parecida, las restricciones $x_j \ge 0$ se conocen como **restricciones de no negatividad** (o *condiciones de no negatividad*).

Otras formas

Debe hacerse notar que el modelo anterior no se ajusta a la forma natural de algunos problemas de programación lineal. Las otras *formas legítimas* son las siguientes:

1. Minimizar en lugar de maximizar la función objetivo:

$$\text{Minimizar} \quad Z = c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

2. Algunas restricciones funcionales con desigualdad en sentido mayor o igual que:

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \ge b_i \quad \text{para algunos valores de } i.$$

3. Algunas restricciones funcionales en forma de ecuación:

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i \quad \text{para algunos valores de } i.$$

4. Algunas variables de decisión sin la restricción de no negatividad:

x_j no está restringida en su signo para algunos valores de j.

Cualquier problema que incluye una, varias o todas estas formas con las otras partes del modelo anterior también se clasifica como un problema de programación lineal. La interpretación de las palabras *asignación de recursos limitados entre actividades* que compiten puede ya no aplicarse

[1] Se llama *nuestra* forma estándar en lugar de *la* forma estándar porque otros libros adoptan formas distintas.

muy bien, si es que se aplica; pero sin importar cuál sea la interpretación o el contexto, lo único necesario es que la formulación matemática del problema se ajuste a las formas permitidas. Así, la definición concisa de un problema de programación lineal es que cada componente de su modelo se ajusta a la forma estándar o a una de las otras formas legítimas que ya se mencionaron.

Terminología de las soluciones del modelo

Puede ser que el lector esté acostumbrado a que el término *solución* signifique la respuesta final a un problema, pero en programación lineal (y sus extensiones) la convención es bastante distinta. Ahora, *cualquier* conjunto de valores específicos de las variables de decisión ($x_1, x_2, \dots, x_n$) se llama una **solución**, aunque sea sólo una posibilidad deseable o ni siquiera permitida. Después se identifican los tipos de soluciones mediante el empleo de un adjetivo apropiado.

Una **solución factible** es aquella para la que *todas* las restricciones se *satisfacen*.
Una **solución no factible** es una solución para la que *al menos una* restricción se *viola*.

En el ejemplo, los puntos (2, 3) y (4, 1) de la figura 3.2 son soluciones factibles, mientras que (−1, 3) y (4, 4) son *soluciones no factibles*.

La **región factible** es la reunión de todas las soluciones factibles.

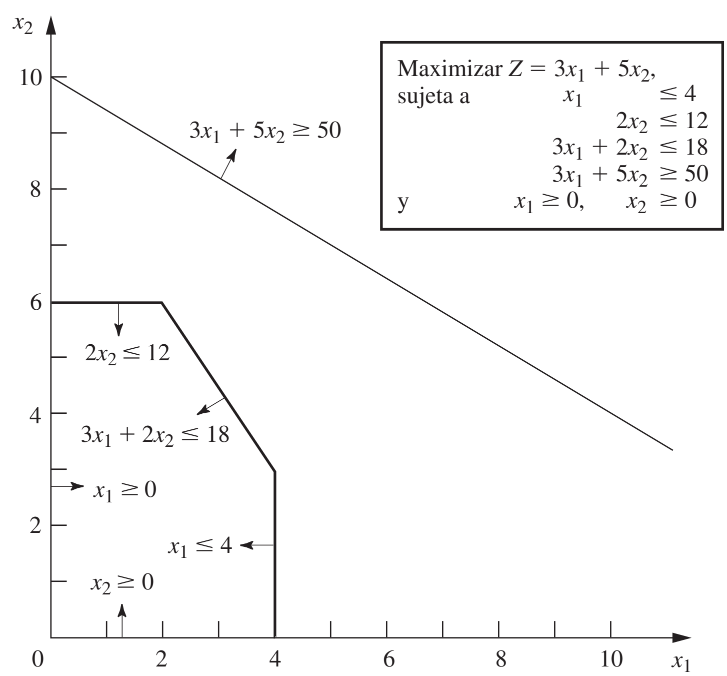
En el ejemplo, la región factible es toda el área sombreada de la figura 3.2.

Es posible que un problema **no tenga soluciones factibles**. Esto habría ocurrido si se hubiera requerido que los nuevos productos tuvieran un rendimiento neto de 50 000 dólares semanales por lo menos, para justificar la interrupción de la fabricación de la línea actual. La restricción correspondiente, $3x_1 + 5x_2 \geq 50$, hubiera eliminado por completo la región factible, con lo que ninguna mezcla de nuevos productos sería superior a la situación actual. Este caso se ilustra en la figura 3.4.

Dado que existen soluciones factibles, la meta de la programación lineal es encontrar una solución factible que sea la mejor, medida por el valor de la función objetivo en el modelo.

Una **solución óptima** es una solución factible que proporciona el *valor más favorable* de la función objetivo.

FIGURA 3.4
El problema de la Wyndor Glass Co. no tendría soluciones óptimas si se le agregara la restricción $3x_1 + 5x_2 \geq 50$.



El **valor más favorable** significa el *valor más grande* si la función objetivo debe *maximizarse*, o el *valor más pequeño* si la función objetivo debe *minimizarse*.

La mayor parte de los problemas tendrá nada más una solución óptima. Sin embargo, también es posible tener más de una. Esto ocurriría en el ejemplo si la *ganancia por lote producido* del producto 2 se cambiara a 2 000 dólares. Este hecho cambiaría la función objetivo a $Z = 3x_1 + 2x_2$, de manera que todos los puntos sobre el segmento de recta que va de (2, 6) a (4, 3) serían soluciones óptimas, situación que se ilustra en la figura 3.5. Igual que en este caso, *cualquier* problema que tenga **soluciones óptimas múltiples** tendrá un número *infinito* de ellas, todas con el mismo valor de la función objetivo.

Otra posibilidad es que el problema **no tenga soluciones óptimas**, lo cual ocurre sólo si: 1) no tiene soluciones factibles, o 2) las restricciones no impiden que el valor de la función objetivo (Z) mejore indefinidamente en la dirección favorable (positiva o negativa). Este caso se conoce como un problema con **Z no acotada** u *objetivo no acotado*. El último caso sería cierto si, por error, en el ejemplo se omitieran las últimas dos restricciones funcionales del modelo, lo cual se ilustra en la figura 3.6.

Ahora se introducirá un tipo especial de soluciones factibles que tiene un papel importante cuando el método símplex trata de encontrar una solución óptima.

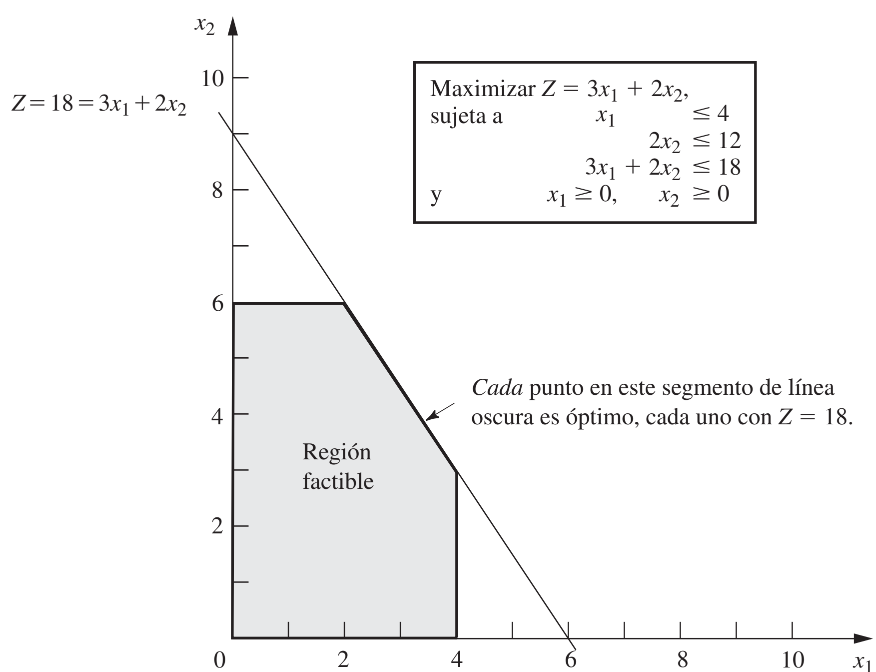
> Una **solución factible en un vértice (FEV)** es una solución que se encuentra en una esquina de la región factible.

(Las soluciones FEV también se conocen como *puntos extremos* o *esquinas*, pero preferimos la terminología más sugerente de *vértice*.) La figura 3.7 pone de relieve cinco soluciones factibles en los vértices del ejemplo.

En las secciones 4.1 y 5.1 se analizarán varias propiedades útiles de las soluciones FEV para problemas de cualquier tamaño, incluso la siguiente relación con las soluciones óptimas.

> **Relación entre las soluciones óptimas y las soluciones FEV:** Considere cualquier problema de programación lineal con soluciones factibles y una región factible acotada. El problema *debe* poseer soluciones FEV y al menos una solución óptima. Además, la mejor solución FEV *debe* ser una solución óptima. Entonces, si un problema tiene exactamente una solución óptima, ésta *debe* ser una solución FEV. Si el problema tiene múltiples soluciones óptimas, al menos dos deben ser soluciones FEV.

■ **FIGURA 3.5**
El problema de la Wyndor Glass Co. tendría múltiples soluciones óptimas si la función objetivo se cambiara a $Z = 3x_1 + 2x_2$.



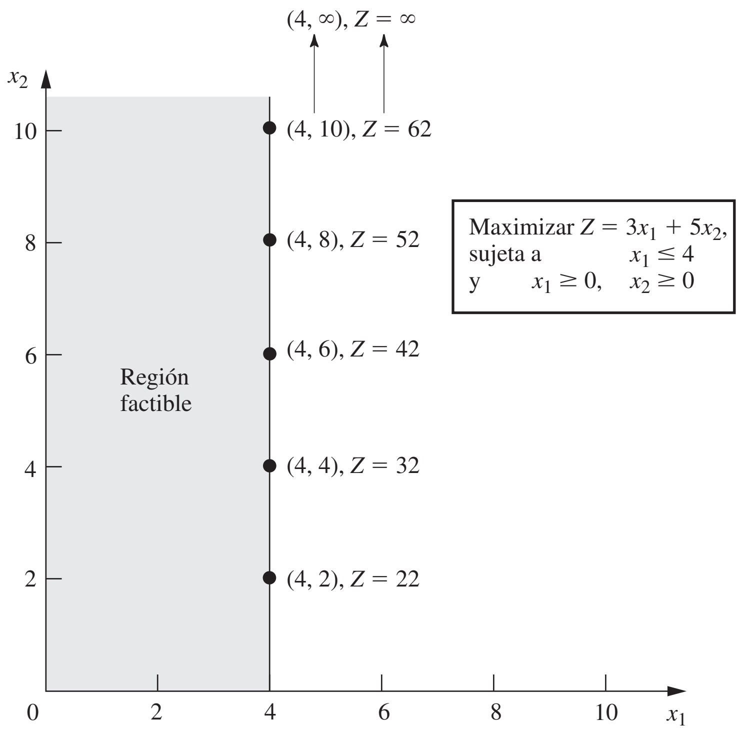


FIGURA 3.6
El problema de la Wyndor Glass Co. no tendría soluciones óptimas si la única restricción funcional fuera $x_1 \leq 4$, puesto que x_2 podría aumentar de modo indefinido en la región factible sin llegar a un valor máximo de $Z = 3x_1 + 5x_2$.

El ejemplo tiene exactamente una solución óptima, $(x_1, x_2) = (2, 6)$, que es FEV. (Considérese la forma como el método gráfico que conduce a la solución óptima que es FEV.) Cuando se modifica el ejemplo para que tenga soluciones óptimas múltiples, como se muestra en la figura 3.5, dos de estas soluciones óptimas —(2, 6) y (4, 3)— son soluciones factibles en los vértices.

3.3 SUPUESTOS DE PROGRAMACIÓN LINEAL

En realidad, todos los supuestos de programación lineal están implícitos en la formulación del modelo que se presentó en la sección 3.2. En particular, desde un punto de vista matemático, los supuestos simplemente son que el modelo debe tener una función objetivo lineal sujeta a restricciones lineales. Sin embargo, desde el punto de vista de modelación, estas propiedades

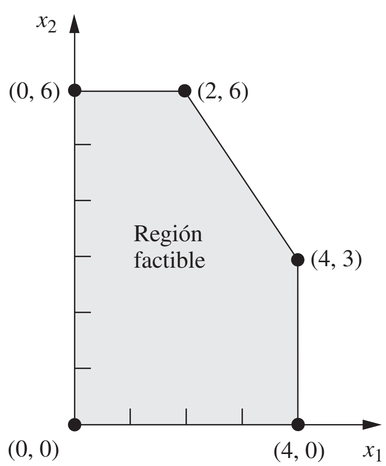


FIGURA 3.7
Los puntos indican las cinco soluciones FEV para el problema de la Wyndor Glass Co.

TABLA 3.4 Ejemplos de proporcionalidad satisfecha o violada

	Ganancia del producto 1 ($000 por semana)			
	Proporcionalidad satisfecha	Proporcionalidad violada		
x_1		Caso 1	Caso 2	Caso 3
0	0	0	0	0
1	3	2	3	3
2	6	5	7	5
3	9	8	12	6
4	12	11	18	6

matemáticas de un modelo de programación lineal implican que se deben considerar ciertos supuestos acerca de las actividades y datos del problema que será modelado, incluso algunos acerca del efecto de las variaciones en el nivel de las actividades. Vale la pena hacer hincapié en ellas para que sea más sencillo evaluar si esta técnica es adecuada para un problema dado. Aún más, es necesario analizar por qué el equipo de IO de la Wyndor Glass Co. concluyó que la formulación de programación lineal proporcionaba una representación satisfactoria del problema.

Proporcionalidad

La *proporcionalidad* es un supuesto sobre la función objetivo y sobre las restricciones funcionales, como se resume a continuación.

> **Supuesto de proporcionalidad:** La contribución de cada actividad al *valor de la función objetivo Z* es *proporcional* al *nivel de la actividad* x_j, como lo representa el término $c_j x_j$ en la función objetivo. De manera similar, la contribución de cada actividad *al lado izquierdo de cada restricción funcional* es *proporcional* al *nivel de la actividad* x_j, como lo representa en la restricción el término $a_{ij} x_j$. En consecuencia, este supuesto elimina cualquier exponente diferente de 1 para las variables en cualquier término de las funciones —ya sea la función objetivo o la función en el lado izquierdo de las restricciones funcionales— en un modelo de programación lineal.[2]

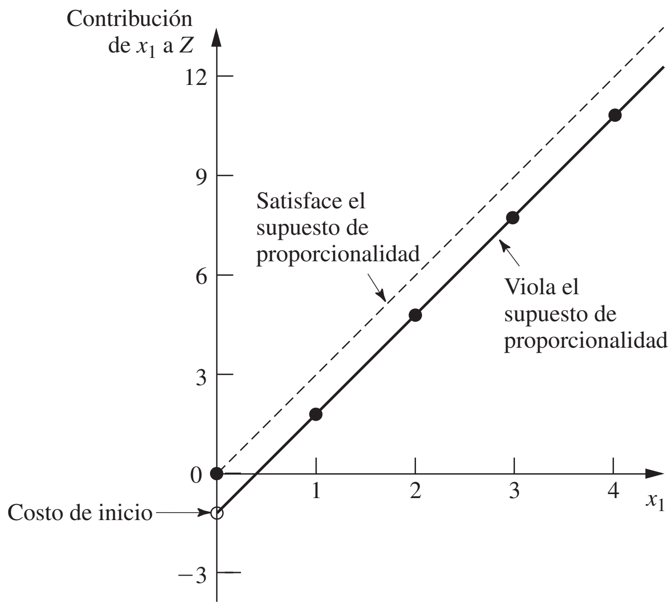
Para ilustrar este supuesto, considere el primer término ($3x_1$) en la función objetivo ($Z = 3x_1 + 5x_2$) del problema de la Wyndor Glass Co. Este término representa la ganancia generada por semana (en miles de dólares) cuando se fabrica el producto 1 a una tasa de x_1 lotes por semana. La columna de *proporcionalidad satisfecha* de la tabla 3.4 muestra el caso que se supuso en la sección 3.1, esto es, que la ganancia sin duda es proporcional a x_1 de manera que $3x_1$ es el término apropiado de la función objetivo. Por el contrario, las siguientes tres columnas muestran casos hipotéticos diferentes en los que el supuesto de proporcionalidad no se cumple.

Vea primero la columna del *caso 1* en la tabla 3.4. Este caso surgiría si se tuvieran *costos fijos* asociados al arranque de la fabricación del producto 1. Por ejemplo, es posible que existan costos debidos a la preparación de las instalaciones de producción. También puede haber costos asociados con el arreglo de la distribución del nuevo producto. Como se trata de costos en los que se incurre una sola vez, deben amortizarse cada semana para que sean conmensurables con Z (ganancia en miles de dólares por semana). Suponga que se hace esta amortización y que los costos de preparación o fijos totales significan una reducción de 1 en el valor de Z, la ganancia, sin considerar los costos fijos es de $3x_1$. Esto quiere decir que la contribución del producto 1 a Z es $3x_1 - 1$ para $x_1 > 0$, mientras que la contribución es $3x_1 = 0$ cuando $x_1 = 0$ (no hay costo fijo). Esta función de ganancias,[3] dada por la curva continua en la figura 3.8, sin duda *no* es proporcional a x_1.

---

[2] Cuando la función incluye *algún término de producto cruzado*, la proporcionalidad debe interpretarse en el sentido de que los *cambios* en el valor de la función son proporcionales a los *cambios* en cada variable (x_j) en forma individual, dados cualesquiera valores fijos para las otras variables. Por lo tanto, un término de producto cruzado satisface la proporcionalidad siempre que cada variable del término tenga un exponente de 1. (Sin embargo, cualquier término de producto cruzado viola el *supuesto de aditividad* que se estudiará en seguida.)

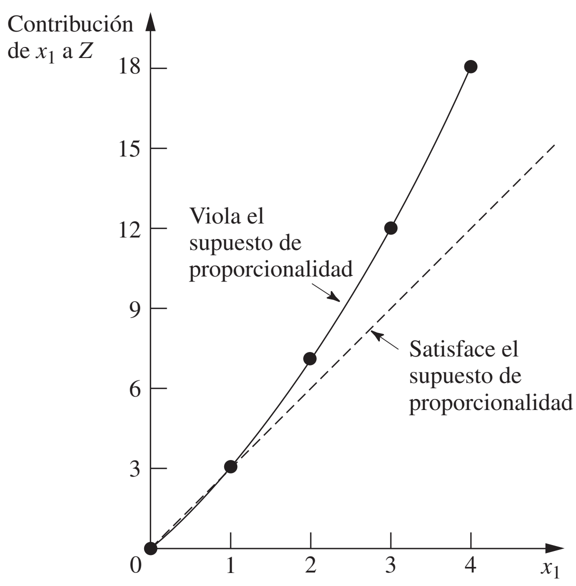
[3] Si la contribución del producto 1 a la función Z fuera $3x_1 - 1$ para *toda* $x_1 \geq 0$, incluso $x_1 = 0$, entonces la constante fija, -1, se podría eliminar de la función objetivo sin cambiar la solución óptima y se restablecería la proporcionalidad. Sin embargo, en este caso no es posible, ya que la constante -1 no se aplica si $x_1 = 0$.

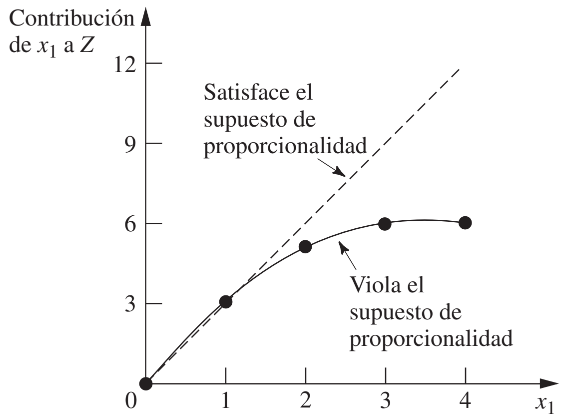


■ **FIGURA 3.8**
La curva continua viola el supuesto de proporcionalidad debido al costo fijo en que se incurre cuando x_1 aumenta desde cero. Los valores de los puntos están dados en la columna del caso 1 de la tabla 3.4.

A primera vista, podría parecer que el *caso 2* de la tabla 3.4 es bastante parecido al caso 1. Pero el hecho es que el caso 2 surge de forma muy diferente. No existe un costo fijo y la ganancia generada por la primera unidad del producto 1 por semana, por supuesto, es de 3 dólares, como se supuso en un principio. Pero ahora se tiene un *rendimiento marginal creciente*; es decir, la *pendiente* de la *función de ganancia* del producto 1 (vea la curva continua de la figura 3.9) crece a medida que x_1 aumenta. Esta violación de la proporcionalidad puede ocurrir debido a economías de escala que en ocasiones se pueden lograr en niveles altos de producción, por ejemplo, a través del uso de maquinaria más eficiente para altos volúmenes, corridas de producción más grandes, descuentos por cantidad por compras grandes de materia prima y por el efecto de la curva de aprendizaje debido a la cual los trabajadores son cada vez más eficientes a medida que adquieren experiencia en un trabajo de producción dado. Cuando el costo incremental disminuye, la ganancia incremental aumenta si se supone un ingreso marginal constante.

■ **FIGURA 3.9**
La curva continua viola el supuesto de proporcionalidad porque su pendiente (el *rendimiento marginal* del producto 1) sigue en crecimiento a medida que x_1 aumenta. Los valores de los puntos están dados en la columna del caso 2 de la tabla 3.4.





■ **FIGURA 3.10**
La curva continua viola el supuesto de proporcionalidad porque su pendiente (el *rendimiento marginal* del producto 1) sigue en decadencia a medida que x_1 aumenta. Los valores de los puntos están dados en la columna del caso 3 de la tabla 3.4.

De nuevo, según los datos de la tabla 3.4, el caso contrario del 2 es el *caso 3*, en el que existe un *rendimiento marginal decreciente*. En este caso, la *pendiente* de la *función de ganancia* del producto 1 (dada por la curva continua de la figura 3.10) disminuye conforme x_1 aumenta. Esta violación de la proporcionalidad puede ocurrir debido a que los *costos de marketing* tienen que elevarse más que proporcionalmente para lograr aumentos del nivel de ventas. Por ejemplo, tal vez el producto 1 se pueda vender a una tasa de 1 por semana ($x_1 = 1$) sin publicidad, mientras que lograr ventas que sostengan una tasa de producción de $x_1 = 2$ puede requerir una publicidad moderada, para $x_1 = 3$ es posible que sea necesaria una extensa campaña publicitaria y para $x_1 = 4$ puede requerirse también una disminución de precios.

Los tres casos son ejemplos hipotéticos de la forma en que el supuesto de proporcionalidad puede no cumplirse. ¿Cuál es la situación real? La ganancia real al fabricar el producto 1 (o cualquier otro) se deriva del ingreso por ventas menos los distintos costos directos e indirectos. Es inevitable que algunos de estos componentes de costos no sean estrictamente proporcionales a las tasas de producción, tal vez por alguna de las razones que se expusieron. Sin embargo, la pregunta importante es si después de acumular todos los componentes de ganancia, la proporcionalidad es una aproximación razonable del modelado. En el problema de la Wyndor Glass Co., el equipo de IO verificó tanto la función objetivo como las restricciones funcionales. La conclusión fue que sin duda podía suponerse la proporcionalidad sin distorsiones serias.

¿Qué ocurre cuando el supuesto no se cumple, ni siquiera como una aproximación razonable? En la mayor parte de los casos, esto significa que se debe emplear *programación no lineal* (vea el capítulo 12). Sin embargo, en la sección 12.8 se señala que cierta clase importante de falta de proporcionalidad se puede manejar mediante programación lineal a través de la reelaboración del problema de manera adecuada. Aún más, si se viola el supuesto nada más debido a los costos fijos, existe una extensión de la programación lineal (*programación entera mixta*) que se puede usar, presentada en la sección 11.3 (el problema de costos fijos).

Aditividad

Aunque el supuesto de proporcionalidad elimina los exponentes diferentes de uno, no prohíbe los *términos de productos cruzados*, términos que incluyen el producto de dos o más variables. El supuesto de aditividad elimina esta posibilidad, como se ve a continuación.

> **Supuesto de aditividad:** *Cada* función de un modelo de programación lineal (ya sea la función objetivo o el lado izquierdo de las restricciones funcionales) es la *suma* de las *contribuciones individuales* de las actividades respectivas.

Para que esta definición sea más concreta y aclare por qué es necesario preocuparse por este supuesto se analizarán algunos ejemplos. En la tabla 3.5 se muestran algunos casos posibles de la función objetivo del problema de la Wyndor Glass Co. En cada caso, las *contribuciones individuales* de los productos son las que se supusieron en la sección 3.1, es decir, $3x_1$ para el producto 1 y $5x_2$ para el producto 2. La diferencia estriba en el último renglón que da el *valor de la función* de

TABLA 3.5 Ejemplos que satisfacen o violan la aditividad de la función objetivo

	Valor de Z		
		Aditividad violada	
(x_1, x_2)	Aditividad satisfecha	Caso 1	Caso 2
(1, 0)	3	3	3
(0, 1)	5	5	5
(1, 1)	8	9	7

Z cuando se fabrican los dos productos de manera conjunta. La columna de *aditividad satisfecha* muestra el caso en el que este *valor de la función* se obtiene simplemente mediante la suma de los dos primeros renglones ($3 + 5 = 8$), es decir, $Z = 3x_1 + 5x_2$, como se supuso antes. Por el contrario, las columnas que siguen muestran casos hipotéticos en los que el supuesto de aditividad queda violado, pero no el de proporcionalidad.

De acuerdo con la columna del *caso 1* de la tabla 3.5, se tiene una función objetivo de $Z = 3x_1 + 5x_2 + x_1x_2$, de manera que $Z = 3 + 5 + 1 = 9$ para $(x_1, x_2) = (1, 1)$, lo que viola el supuesto de aditividad de que $Z = 3 + 5$. (El supuesto de proporcionalidad todavía se satisface puesto que si se fija el valor de una variable, el incremento de Z debido a la otra variable es proporcional al valor de esa variable.) Este caso surge si los dos productos son *complementarios* de alguna forma en que la ganancia *aumenta*. Por ejemplo, suponga que es necesaria una campaña publicitaria importante para comercializar cualquiera de los dos productos por sí solos, pero que la misma campaña puede promover de manera eficaz ambos productos. Como se ahorra un costo alto del segundo producto, la ganancia conjunta será algo más que la *suma* de sus ganancias individuales si se producen por separado.

El *caso 2* de la tabla 3.5 también viola el supuesto de aditividad debido al término adicional en su función objetivo, $Z = 3x_1 + 5x_2 - x_1x_2$, de forma que $Z = 3 + 5 - 1 = 7$ para $(x_1, x_2) = (1, 1)$. Al contrario del primer caso, el caso 2 surge cuando los dos productos son *competitivos* de algún modo en que su ganancia conjunta *disminuye*. Por ejemplo, suponga que ambos productos deben usar la misma maquinaria y equipo. Si se produce cada uno por sí solo, maquinaria y equipo se dedican a este único uso. Sin embargo, cuando son fabricados ambos productos, se requiere cambiar los procesos de producción de uno a otro, con tiempo y costos involucrados en la interrupción temporal de la producción de uno y la preparación del otro. Debido a este costo adicional importante, su ganancia conjunta será algo menor que la *suma* de sus ganancias individuales si se les produce por separado.

El mismo tipo de interacción de actividades puede afectar la aditividad de las funciones de restricción. Por ejemplo, considere la tercera restricción del problema de la Wyndor Glass Co., $3x_1 + 2x_2 \leq 18$. (Ésta es la única restricción que incluye ambos productos.) Esta restricción se refiere a la capacidad de producción de la planta 3, en la que se dispone de 18 horas semanales de producción para los dos nuevos productos, mientras que la función del lado izquierdo ($3x_1 + 2x_2$) representa el número de horas de producción semanales que se usarían en estos productos. La columna de *aditividad satisfecha* de la tabla 3.6 muestra este caso, mientras que las dos columnas siguientes exponen casos en los que la función tiene un término adicional de producto cruzado que

TABLA 3.6 Ejemplos de aditividad satisfecha o violada de una restricción funcional

	Cantidad de recurso usado		
		Aditividad satisfecha	
(x_1, x_2)	Aditividad violada	Caso 3	Caso 4
(2, 0)	6	6	6
(0, 3)	6	6	6
(2, 3)	12	15	10.8

viola la aditividad. En las tres columnas, las *contribuciones individuales* de los productos en cuanto al uso de la capacidad de la planta 3 son las que se supusieron, es decir, $3x_1$ para el producto 1 y $2x_2$ para el producto 2, o sea, $3(2) = 6$, para $x_1 = 2$ y $2(3) = 6$ para $x_2 = 3$. Igual que en la tabla 3.5, la diferencia estriba en el último renglón que ahora da el *valor total de la función* para el tiempo de producción que se utiliza cuando se fabrican los dos productos de manera conjunta.

En el caso 3 (vea la tabla 3.6), el tiempo de producción para los dos productos está dado por la función $3x_1 + 2x_2 + 0.5x_1x_2$, de manera que el *valor total de la función* es $6 + 6 + 3 = 15$ cuando $(x_1, x_2) = (2, 3)$, lo que viola el supuesto de aditividad de que el valor es sólo $6 + 6 = 12$. Este caso puede surgir justo de la misma forma que se describió en el caso 2 (tabla 3.5): tiempo adicional desperdiciado en el cambio de procesos de producción entre los dos productos. El término adicional de producto cruzado ($0.5x_1x_2$) representa el tiempo de producción desperdiciado en esta forma. (Observe que el desperdicio de tiempo al cambiar de un producto a otro da por resultado, en este caso, un término positivo de producto cruzado en donde la función total mide el tiempo de producción utilizado, mientras que lleva a un término negativo de producto cruzado en el caso 2 puesto que esa función total mide la ganancia.)

En el caso 4 de la tabla 3.6, la función de la capacidad que se usa es $3x_1 + 2x_2 - 0.1x_1^2x_2$, por lo que el *valor de la función* para $(x_1, x_2) = (2, 3)$ es $6 + 6 - 1.2 = 10.8$. Este caso surge de la siguiente manera. Igual que en el caso 3, suponga que los dos productos requieren el mismo tipo de maquinaria y equipo, pero que ahora el tiempo para cambiar de un producto a otro es relativamente pequeño. Como cada producto pasa por una serie de operaciones, las instalaciones de producción individual, que por lo general se dedican a ese producto, tendrían algunos tiempos ociosos. Estas instalaciones podrían utilizarse durante estos tiempos en otros productos. En consecuencia, el tiempo total de producción usado cuando se fabrican en forma conjunta los dos productos, es menor que la *suma* de los tiempos de producción usados por los productos individuales cuando se fabrican por separado.

Después de analizar los tipos posibles de interacción de los dos productos ilustrados en estos cuatro casos, el equipo de IO concluyó que ninguno tenía un papel importante en el problema real de la Wyndor Glass Co. Por lo tanto, el supuesto de aditividad se adoptó como una aproximación razonable.

En otros problemas, si la aditividad no es un supuesto razonable, de forma que algunas o todas las funciones matemáticas del modelo necesariamente son *no lineales* (debido a términos de producto cruzado), resulta definitiva la entrada en el ámbito de la programación no lineal (capítulo 12).

Divisibilidad

El siguiente supuesto se refiere a los valores permitidos para las variables de decisión.

> **Supuesto de divisibilidad:** En un modelo de programación lineal, las variables de decisión pueden tomar *cualquier* valor, incluso valores *no enteros*, que satisfagan las restricciones funcionales y de no negatividad. En consecuencia, estas variables *no* están restringidas a sólo valores enteros. Como cada variable de decisión representa el nivel de alguna actividad, se supondrá que las actividades se pueden realizar a *niveles fraccionales*.

En el problema de la Wyndor Glass Co., las variables de decisión representan tasas de producción (número de lotes de un producto fabricados a la semana). Como estas tasas pueden tomar *cualquier* valor fraccional dentro de la región factible, el supuesto de divisibilidad se cumple.

En ciertas situaciones, el supuesto de divisibilidad no se cumple porque algunas o todas las variables de decisión deben restringirse a *valores enteros*. Los modelos matemáticos con esta restricción, que se llaman modelos de *programación entera*, se estudiarán en el capítulo 11.

Certidumbre

El último supuesto se refiere a los *parámetros* del modelo, es decir, a los coeficientes c_j, en la función objetivo, los coeficientes a_{ij}, en las restricciones funcionales y los b_i en el lado derecho de las restricciones funcionales.

> **Supuesto de certidumbre:** Se supone que los valores asignados a cada parámetro de un modelo de programación lineal son *constantes conocidas*.

En los problemas reales, el supuesto de certidumbre casi nunca se satisface por completo. Por lo general se formula un modelo de programación lineal para elegir un curso de acción futuro. En este caso, los valores de los parámetros que se emplean están basados en una predicción de las condiciones futuras, lo que es inevitable que introduzca cierto grado de incertidumbre.

Por esta razón, siempre es importante realizar un **análisis de sensibilidad** después de encontrar una solución óptima de los valores supuestos de los parámetros. Como se presentó en la sección 2.3, el propósito general es identificar los parámetros *sensibles* (es decir, aquellos cuyo valor no puede cambiar mucho sin cambiar la solución óptima), debido a que un cambio mayor en el valor de un parámetro sensible de inmediato envía la señal de la necesidad de introducir un cambio en la solución usada.

El análisis de sensibilidad tiene un papel importante en el problema de la Wyndor Glass Co., como se verá en la sección 6.7. De cualquier manera, es necesario adquirir algunos conocimientos adicionales antes de terminar esta historia.

En algunos casos, el grado de incertidumbre en los parámetros es demasiado grande para que el análisis de sensibilidad lo pueda manejar. En estas situaciones es necesario establecer, en forma explícita, estos parámetros como *variables aleatorias*. Se han desarrollado formulaciones de este tipo, que se pueden consultar en las secciones 23.6 y 23.7 del sitio web del libro.

Los supuestos en perspectiva

En la sección 2.2 se hizo hincapié en que el modelo matemático intenta ser sólo una representación idealizada del problema real. Por lo general se requieren aproximaciones y los supuestos de simplificación para que el modelo se pueda manejar. Agregar demasiados detalles y precisión puede hacer que el modelo sea difícil de manipular para llevar a cabo un análisis útil del problema. En realidad, todo lo que se necesita es que exista una correlación relativamente alta entre la predicción del modelo y lo que de hecho pasaría en el problema real.

Este consejo sin duda es aplicable a la programación lineal. Es muy frecuente en las aplicaciones reales de esta técnica que casi *ninguno* de los cuatro supuestos se cumpla. Excepto, quizá, en el caso del *supuesto de divisibilidad*, deben esperarse pequeñas disparidades. Esto es cierto en especial para el *supuesto de certidumbre*, de manera que es normal que deba aplicarse el análisis de sensibilidad para compensar la violación de este supuesto.

Sin embargo, es importante que el equipo de IO examine los cuatro supuestos en el problema que se estudia y analice el tamaño de las disparidades. Si cualquiera de los supuestos es violado de manera importante, es necesario disponer de varios modelos alternativos, como se verá en capítulos posteriores de este libro. Una desventaja de estos modelos es que los algoritmos disponibles para resolverlos no son tan poderosos como el de programación lineal, pero en algunos casos este inconveniente se ha solucionado. En algunas aplicaciones se utiliza el poderoso enfoque de programación lineal para el análisis inicial y después un modelo más complejo para perfeccionar el análisis.

Al trabajar los ejemplos de la siguiente sección se demostrará que el análisis del grado en que se cumplen los cuatro supuestos de la programación lineal es una buena práctica.

3.4 EJEMPLOS ADICIONALES

El problema de la Wyndor Glass Co. es un ejemplo prototípico de programación lineal en varios aspectos: comprende la asignación de recursos limitados entre actividades que compiten por ellos, su modelo se ajusta a la forma estándar y su contexto es el tradicional de planeación para mejorar la administración. Sin embargo, la aplicación de la programación lineal es mucho más extensa. Esta sección comienza por ampliar el horizonte. Al estudiar los siguientes ejemplos observe que se caracterizan como problemas de programación lineal por el modelo matemático, más que por su contexto. Luego, debe considerarse que el mismo modelo matemático surge en muchos otros contextos con sólo cambiar los nombres de las actividades.

Estos ejemplos son versiones simplificadas de aplicaciones reales. Como el problema de Wyndor y el ejemplo de demostración del problema gráfico en el OR Tutor, el primero de estos ejemplos tiene sólo dos variables de decisión, de manera que puede ser resuelto mediante el método gráfico. Ahora se trata de un problema de minimización y tiene una mezcla de formas para

las restricciones funcionales. (Este ejemplo simplifica de manera considerable la situación real de diseñar una terapia de radiación, pero el primer Recuadro de aplicación en esta sección describe el emocionante impacto que en realidad está teniendo la IO en esta área.) Los ejemplos subsecuentes tienen muchas más de dos variables de decisión y por lo tanto son más difíciles de formular. Aunque se mencionarán las soluciones óptimas que se obtienen por medio del método símplex, en esta sección el enfoque se concentra en la manera de formular el modelo de programación lineal para estos problemas más grandes. En las secciones subsecuentes y en el capítulo siguiente se dará mayor importancia a las herramientas de software y al algoritmo (método símplex) que se utiliza para resolver dichos problemas.

Si el lector considera que requiere **ejemplos adicionales** de formulación de modelos de programación lineal pequeños y relativamente directos antes de tratar con los ejemplos de formulación más grandes, se le sugiere regresar al caso de demostración del método gráfico en el OR Tutor y a los ejemplos en la sección de Worked Examples de este capítulo en el sitio web del libro.

Diseño de terapia de radiación

Acaban de diagnosticar que Mary padece cáncer en una etapa bastante avanzada. Específicamente, tiene un tumor grande en el área de la vejiga, una "lesión que afecta a toda la vejiga".

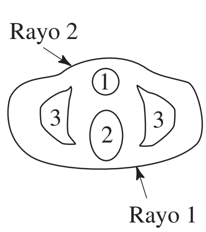
Mary recibirá los cuidados médicos más avanzados disponibles, para proporcionarle la mejor posibilidad de supervivencia. Estos cuidados incluyen una terapia de *radiación extensa*.

La terapia implica el uso de una máquina de rayos externos que envía radiación ionizante a través del cuerpo de la paciente y daña tanto los tejidos cancerosos como los sanos. Es normal que se administren los rayos con precisión desde diferentes ángulos en un plano de dos dimensiones. Debido a la atenuación, cada rayo descarga más radiación sobre el tejido cercano al punto de entrada que sobre el cercano al punto de salida. La dispersión también provoca que parte de la radiación se descargue sobre tejidos que están fuera de la trayectoria directa del rayo. Debido a que las células del tumor casi siempre se encuentran diseminadas entre células sanas, la dosis de radiación a través de la región del tumor debe ser suficiente para matar las células malignas que son un poco más sensibles a ella, pero suficientemente pequeña para no matar a las células sanas. Al mismo tiempo, la dosis acumulada que reciben los tejidos críticos no debe exceder los niveles de tolerancia establecidos, con el objeto de prevenir complicaciones que puedan resultar más serias que la enfermedad misma. La dosis completa que recibe el cuerpo sano debe minimizarse.

Debido a la necesidad de balancear con cuidado todos estos factores, el diseño de la terapia de radiación es un proceso muy delicado. La meta principal de este diseño es elegir la combinación de rayos que se utilizará y la intensidad de cada uno para generar la mejor distribución posible de la dosis. (La fuerza de la dosis en cualquier punto del cuerpo se mide en unidades llamadas *kilorads*.) Una vez diseñado el tratamiento, se administra en muchas sesiones durante varias semanas.

En el caso de Mary, el tamaño y la localización del tumor hacen que el diseño de su tratamiento sea un proceso más delicado que lo usual. La figura 3.11 muestra un diagrama de un corte transversal del tumor visto desde arriba, al igual que los tejidos cercanos críticos que deben evitarse. Estos tejidos incluyen órganos vitales —por ejemplo, el recto— y estructura ósea —el fémur y la pelvis— que atenuarán la radiación. Además, se muestra el punto de entrada y la dirección de los únicos dos rayos que se pueden usar en este caso con un grado relativamente moderado de seguridad. (El ejemplo se ha simplificado en este punto, pero en la realidad se consideran docenas de rayos posibles.)

FIGURA 3.11
Corte transversal del tumor de Mary (visto desde arriba), cerca de tejidos críticos y de los rayos de radiación usados.



1. Vejiga y tumor
2. Recto, cóccix, etc.
3. Fémur, parte de la pelvis, etc.

En el caso de cualquier rayo propuesto de una cierta intensidad, el análisis para determinar cuál sería la absorción de radiación resultante por distintas partes del cuerpo requiere desarrollar un difícil proceso. En resumen, con base en un análisis anatómico cuidadoso, la distribución de energía dentro de un corte transversal de dos dimensiones se puede graficar en un mapa de isodosis en el que las curvas representan la fuerza de la dosis como un porcentaje de la fuerza de ésta en el punto de entrada. Después, se coloca una red fina sobre dicho mapa. Si se suma la radiación absorbida en los cuadros que contienen cada tipo de tejido, se puede calcular la dosis promedio que absorbe el tumor, los tejidos sanos y los tejidos críticos. La absorción de la radiación es aditiva cuando se administra más de un rayo (en forma secuencial).

Después de un análisis exhaustivo, el equipo médico estimó con detalle los datos necesarios para el diseño del tratamiento de Mary, cuyo resumen se presenta en la tabla 3.7. La primera columna presenta una lista de las áreas del cuerpo que deben considerarse y las dos siguientes pro-

Recuadro de aplicación

El *cáncer de próstata* es la forma más común de cáncer diagnosticada en hombres. Se estima que en 2007 hubo 220 000 nuevos casos sólo en Estados Unidos. Como muchas otras formas de esta enfermedad, la *terapia de radiación* es un método común de tratamiento para el cáncer de próstata, cuya meta es tener una dosis de radiación suficientemente alta en la región del tumor para matar las células malignas al mismo tiempo que se minimiza la exposición a la radiación de estructuras sanas críticas cercanas al tumor. Este tratamiento puede aplicarse a través de una terapia de radiación de *rayo externo* (como se ilustra en el primer ejemplo en esta sección) o *radioterapia*, la cual implica colocar alrededor de 100 "semillas" radiactivas dentro de la región del tumor. El reto consiste en determinar el patrón geométrico tridimensional más eficaz para colocarlas.

El **Memorial Sloan-Kettering Cancer Center (MSKCC)** en Nueva York es el centro para el tratamiento del cáncer más antiguo del mundo. Un equipo de IO del *Centro para la Investigación de Operaciones en Medicina y Cuidado de la Salud* del Georgia Institute of Technology trabajó con médicos del MSKCC para desarrollar un *método vanguardista* altamente complejo para optimizar la aplicación de radioterapia al cáncer de próstata. El modelo subyacente se ajusta a la estructura para programación lineal con una excepción. Además de tener las variables continuas comunes que se ajustan a la programación lineal, el modelo también incluye algunas *variables binarias* (variables cuyos únicos valores posibles son 0 y 1). (Este tipo de extensión de la programación lineal a la cual se le llama *programación entera mixta* se explicará en el capítulo 11.) La optimización se hace en cuestión de minutos mediante un sistema de planeación computarizado que puede ser operado con facilidad por el personal médico cuando comienza el procedimiento de insertar las semillas en la próstata del paciente.

Este procedimiento para optimizar la aplicación de la radioterapia para el cáncer de próstata ha teniendo un efecto profundo tanto en los costos del cuidado de la salud como en la calidad de vida de los pacientes que reciben el tratamiento debido a su alta eficacia y a la reducción sustancial de los efectos colaterales. Se estima que cuando todas las clínicas de Estados Unidos adopten este procedimiento, los ahorros en costos anuales serán de alrededor de **500 millones de dólares**, debido a la eliminación de la necesidad de una reunión para planear el tratamiento y una exploración posterior a la operación, así como porque proporciona un procedimiento quirúrgico más eficiente y reduce la necesidad de tratar efectos colaterales subsecuentes. También se anticipa que este enfoque puede extenderse a otras formas de radioterapia, como el tratamiento del cáncer de seno, cérvix, esófago, tracto biliar, páncreas, cabeza y cuello, y ojo.

Esta aplicación de la programación lineal y sus extensiones llevaron al equipo de IO a ganar en 2007 el primer lugar en la competencia internacional por el Premio Franz Edelman al desempeño en investigación de operaciones y ciencias de la administración.

Fuente: E. K. Lee y M. Zaider, "Operations Research Advances Cancer Therapeutics", en *Interfaces*, **38**(1): 5-25, enero-febrero, 2008. (En nuestro sitio web, www.mhhe.com/hillier, se proporciona un vínculo con este artículo.)

porcionan la fracción de la dosis de radiación de cada rayo en el punto de entrada que se absorbe en promedio en las áreas respectivas. Por ejemplo, si el nivel de la dosis en el punto de entrada del rayo 1 es 1 kilorad, se absorberán 0.4 kilorad en toda la anatomía sana en el plano de dos dimensiones, un promedio de 0.3 kilorad en los tejidos críticos cercanos, un promedio de 0.5 kilorad en las distintas partes del tumor *y* 0.6 kilorad en el centro del tumor. La última columna presenta las restricciones sobre la dosis total de ambos rayos que se absorbe en promedio en las diferentes partes del cuerpo. En particular, la absorción promedio de la dosis por la anatomía sana debe ser *tan pequeña como sea posible*, los tejidos críticos no deben exceder 2.7 kilorads, el promedio sobre todo el tumor debe ser *igual* a 6 kilorads y en el centro del tumor debe ser *por lo menos* de 6 kilorads.

TABLA 3.7 Datos para el diseño del tratamiento de radiación de Mary

Área	Fracción de la dosis de entrada absorbida por área (promedio)		Restricción sobre la dosis promedio total, kilorads
	Rayo 1	Rayo 2	
Anatomía sana	0.4	0.5	Minimizar
Tejido crítico	0.3	0.1	$\leq$ 2.7
Región del tumor	0.5	0.5	$=$ 6
Centro del tumor	0.6	0.4	$\geq$ 6

Formulación como un problema de programación lineal. Las decisiones que deben hacerse son las dosis de radiación en los dos puntos de entrada. Por lo tanto, las dos variables de decisión x_1 y x_2 representan la dosis (en kilorads) en el punto de entrada de los rayos 1 y 2, respectivamente. Como debe minimizarse la dosis total que llega a la anatomía sana, se definirá como Z a esta cantidad. En este punto se pueden usar los datos de la tabla 3.7 para formular el siguiente modelo de programación lineal.[4]

$$\text{Minimizar} \quad Z = 0.4x_1 + 0.5x_2,$$

sujeta a

$$0.3x_1 + 0.1x_2 \leq 2.7$$
$$0.5x_1 + 0.5x_2 = 6$$
$$0.6x_1 + 0.4x_2 \geq 6$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

Observe las diferencias entre este modelo y el que se presentó en la sección 3.1 para la Wyndor Glass Co. Este último involucraba *maximizar Z*, y todas las restricciones funcionales tenían la forma $\leq$. El nuevo modelo incorpora otras tres formas *legítimas* descritas en la sección 3.2; a saber: *minimizar Z*, restricciones funcionales de la forma $=$, y restricciones funcionales de la forma $\geq$.

Sin embargo, ambos modelos tienen sólo dos variables, de manera que este nuevo problema también se puede resolver por el *método gráfico* que se ilustró en la sección 3.1. La figura 3.12 muestra la solución gráfica. La *región factible* consiste nada más en el segmento entre los puntos (6, 6) y (7.5, 4.5), ya que los puntos en este segmento son los únicos que satisfacen todas las restricciones al mismo tiempo. (Observe que la restricción de igualdad limita la región factible a la recta que contiene este segmento y las otras restricciones funcionales determinan los puntos extremos del segmento.) La línea punteada representa la función objetivo que pasa por la solución óptima $(x_1, x_2) = (7.5, 4.5)$ con $Z = 5.25$. Esta solución es óptima y no (6, 6) porque *disminuir Z* (para valores positivos de Z) empuja la función objetivo hacia el origen (donde $Z = 0$). Y $Z = 5.25$ para (7.5, 4.5) es menor que $Z = 5.4$ para (6, 6).

En consecuencia, el diseño óptimo implica utilizar una dosis total en el punto de entrada de 7.5 kilorads para el rayo 1 y 4.5 kilorads para el rayo 2.

Planeación regional

La CONFEDERACIÓN SUR DE KIBBUTZIM está formada por tres kibbutzim (comunidades agrícolas comunales) de Israel. La planeación global de este grupo se hace en su oficina de coordinación técnica. En la actualidad planean la producción agrícola para el año próximo.

La producción agrícola está limitada tanto por la extensión de terreno disponible para irrigación como por la cantidad de agua que la Comisión de Aguas (una oficina del gobierno nacional) asigna para irrigarlo. La tabla 3.8 contiene los datos.

Los tipos de cultivos adecuados para la región incluyen remolacha, algodón y sorgo, que son precisamente los tres que están en estudio para la estación venidera. Los cultivos difieren primordialmente en su rendimiento neto esperado por acre y en su consumo de agua. Además, el Ministerio de Agricultura ha establecido una cantidad máxima de acres que la Confederación puede dedicar a estos cultivos. La tabla 3.9 muestra estas cantidades.

[4] Este modelo es *mucho* más pequeño del que normalmente se necesitaría para aplicaciones reales. Para obtener mejores resultados, un modelo realista podría incluso requerir muchas decenas de miles de variables de decisión y restricciones. Por ejemplo, vea H. E. Romeijn, R. K. Ahuja, J. F. Dempsey y A. Kumar, "A New Linear Programming Approach to Radiation Therapy Treatment Planning Problems", en *Operations Research*, **54**(2): 201-216, marzo-abril de 2006. Para enfoques alternativos que combinan programación lineal con otras técnicas de IO (como en el primer Recuadro de aplicación de esta sección), también vea G. J. Lim, M. C. Ferris, S. J. Wright, D. M. Shepard y M. A. Earl, "An Optimization Framework for Conformal Radiation Treatment Planning", en *INFORMS Journal on Computing*, **19**(3): 366-380, verano del 2007.

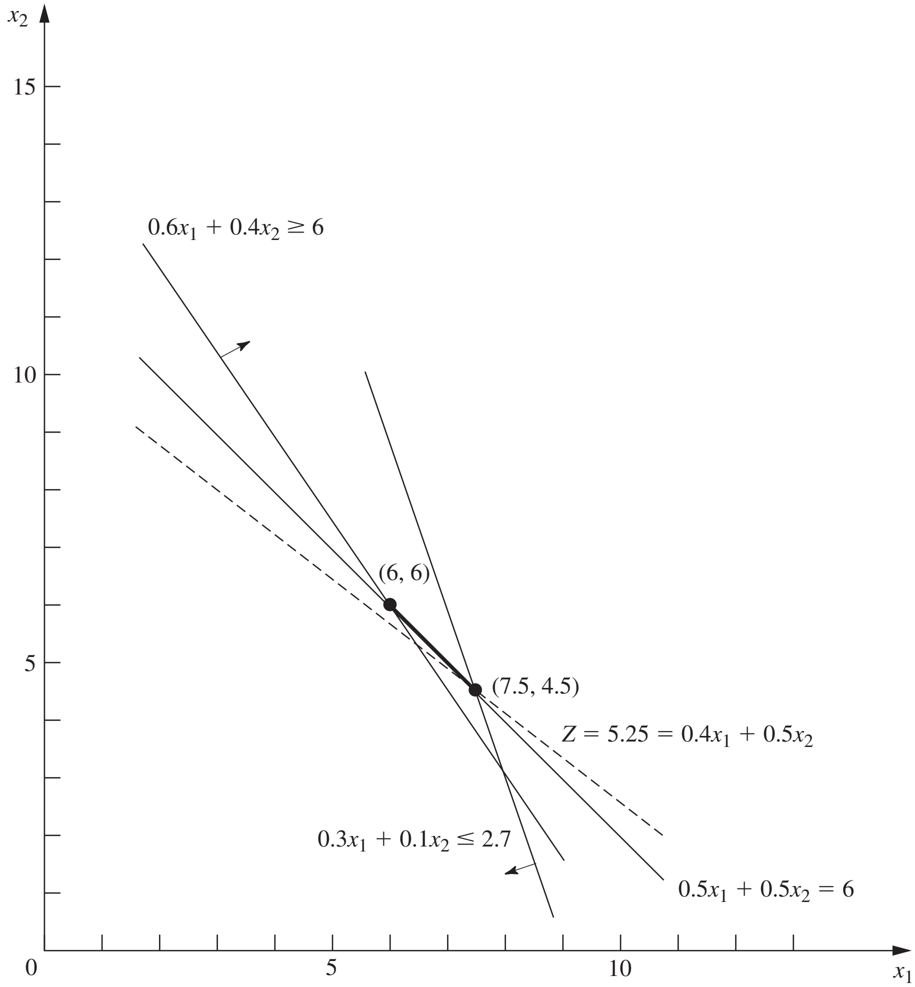


FIGURA 3.12
Solución gráfica del diseño de la terapia de radiación de Mary.

Debido a la disponibilidad limitada de agua para irrigación, la Confederación no podrá usar todo el terreno irrigable para los cultivos de la próxima temporada. Para asegurar la equidad entre los tres kibbutzim, han acordado que cada uno sembrará la misma proporción de sus tierras irrigables disponibles. Por ejemplo, si el kibbutz 1 siembra 200 de sus 400 acres disponibles, entonces el kibbutz 2 deberá sembrar 300 de sus 600 acres, mientras que el kibbutz 3 sembraría 150 acres de los 300 que tiene. Cualquier combinación de estos cultivos se puede sembrar en cualquiera de las granjas. El trabajo al que se enfrenta la oficina de coordinación técnica consiste en planear cuántos acres deben asignarse a cada tipo de cultivo en cada kibbutz, de forma que cumpla con las restricciones dadas. El objetivo es maximizar el rendimiento neto total de la Confederación Sur de Kibbutzim.

Formulación como un problema de programación lineal. Las cantidades sobre las que se tomará la decisión son el número de acres que se dedicará a cada cultivo en cada kibbutz.

TABLA 3.8 Datos de recursos de la Confederación Sur de Kibbutzim

Kibbutz	Terreno disponible (acres)	Asignación de agua (pies-acre)
1	400	600
2	600	800
3	300	375

TABLA 3.9 Datos de cultivos de la Confederación Sur de Kibbutzim

Cultivo	Cantidad máxima (acres)	Consumo de agua (acre-pie/acre)	Rendimiento neto ($/acre)
Remolacha	600	3	1 000
Algodón	500	2	750
Sorgo	325	1	250

Las variables de decisión x_j ($j = 1, 2, \ldots, 9$) representan estas nueve cantidades, como se muestra en la tabla 3.10.

Como la medida de eficacia Z es el rendimiento neto total, el modelo de programación lineal que resulta para este problema es

$$\text{Maximizar} \quad Z = 1\,000(x_1 + x_2 + x_3) + 750(x_4 + x_5 + x_6) + 250(x_7 + x_8 + x_9),$$

sujeta a las siguientes restricciones:

1. Terreno para uso en cada kibbutz:

$$x_1 + x_4 + x_7 \leq 400$$
$$x_2 + x_5 + x_8 \leq 600$$
$$x_3 + x_6 + x_9 \leq 300$$

2. Asignación de agua para cada kibbutz:

$$3x_1 + 2x_4 + x_7 \leq 600$$
$$3x_2 + 2x_5 + x_8 \leq 800$$
$$3x_3 + 2x_6 + x_9 \leq 375$$

3. Total de acres para cada cultivo:

$$x_1 + x_2 + x_3 \leq 600$$
$$x_4 + x_5 + x_6 \leq 500$$
$$x_7 + x_8 + x_9 \leq 325$$

4. Igual proporción de área plantada:

$$\frac{x_1 + x_4 + x_7}{400} = \frac{x_2 + x_5 + x_8}{600}$$
$$\frac{x_2 + x_5 + x_8}{600} = \frac{x_3 + x_6 + x_9}{300}$$
$$\frac{x_3 + x_6 + x_9}{300} = \frac{x_1 + x_4 + x_7}{400}$$

TABLA 3.10 Variables de decisión del problema de la Confederación Sur de Kibbutzim

	Asignación (acres)		
	Kibbutz		
Cultivo	**1**	**2**	**3**
Remolacha	x_1	x_2	x_3
Algodón	x_4	x_5	x_6
Sorgo	x_7	x_8	x_9

TABLA 3.11 Solución óptima del problema de la Confederación Sur de Kibbutzim

	Mejor asignación (acres)		
	Kibbutz		
Cultivo	**1**	**2**	**3**
Remolacha	$133\frac{1}{3}$	100	25
Algodón	100	250	150
Sorgo	0	0	0

5. No negatividad:

$$x_j \geq 0, \quad \text{para } j = 1, 2, \ldots, 9.$$

Esto completa el modelo, a excepción de las igualdades que no están en la forma apropiada para un modelo de programación lineal porque algunas variables están en el lado derecho de las ecuaciones. En consecuencia, la forma final[5] es

$$3(x_1 + x_4 + x_7) - 2(x_2 + x_5 + x_8) = 0$$
$$(x_2 + x_5 + x_8) - 2(x_3 + x_6 + x_9) = 0$$
$$4(x_3 + x_6 + x_9) - 3(x_1 + x_4 + x_7) = 0$$

La oficina de coordinación técnica formuló este modelo y después aplicó el método símplex (que se desarrolla en el capítulo 4) para encontrar una solución óptima

$$(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9) = \left(133\frac{1}{3}, 100, 25, 100, 250, 150, 0, 0, 0\right),$$

como se muestra en la tabla 3.11. El valor óptimo de la función objetivo que obtuvieron es $Z = 633, 333\frac{1}{3}$, es decir, un rendimiento neto total de \$633 333.33.

Control de la contaminación del aire

La NORI & LEETS CO., una de las mayores productoras de acero del mundo occidental, está localizada en la ciudad de Steeltown y es la única empresa grande de la localidad. La comunidad ha crecido y prosperado junto con la compañía, que de momento emplea cerca de 50 000 residentes. La actitud de los habitantes ha sido siempre "lo que es bueno para Nori & Leets es bueno para nosotros". Sin embargo, esta actitud está cambiando; la contaminación no controlada del aire debida a los altos hornos de la planta está en camino de arruinar la apariencia de la ciudad y de poner en peligro la salud de sus habitantes.

Como resultado, después de una revuelta entre los accionistas se eligió un nuevo consejo directivo más responsable. Los nuevos directores han decidido seguir políticas de responsabilidad social y realizar pláticas con las autoridades de la ciudad y con grupos de ciudadanos para tomar medidas respecto de la contaminación ambiental. Juntos han establecido estándares rigurosos de calidad del aire para la ciudad de Steeltown.

Los tres tipos principales de contaminantes son partículas de materia, óxidos de azufre e hidrocarburos. Los nuevos estándares requieren que la compañía reduzca su emisión anual de estos contaminantes en las cantidades que se presentan en la tabla 3.12. El consejo directivo ha dado

[5] En realidad, cualquiera de estas ecuaciones es redundante y se puede eliminar si así se desea. Debido a la forma de las ecuaciones, cualesquiera dos de las restricciones referentes al terreno útil también se pueden eliminar porque se satisfacen de manera automática cuando se satisfacen las restricciones restantes del terreno útil y estas ecuaciones. Sin embargo, la inclusión de restricciones no necesarias no produce problemas (excepto un esfuerzo computacional un poco mayor), por lo que no es necesario preocuparse por identificarlas y eliminarlas del modelo formulado.

TABLA 3.12 Estándares de aire limpio de Nori & Leets Co.

Contaminante	Reducción requerida de la tasa de emisión anual (millones de libras)
Partículas	60
Óxidos de azufre	150
Hidrocarbonos	125

instrucciones a la administración para que el personal de ingeniería determine cómo lograr estas reducciones en la forma más económica.

La fabricación de acero tiene dos fuentes principales de contaminación: los altos hornos para fabricar el arrabio (lingotes de hierro) y los hornos Siemens-Martin para transformar el hierro en acero. En ambos casos, los ingenieros determinaron que los métodos de abatimiento más eficaces son: 1) aumentar la altura de las chimeneas,[6] 2) usar filtros (con trampas de gas) en ellas y 3) incluir limpiadores de alto grado en los combustibles de los hornos. Todos estos métodos tienen limitaciones tecnológicas en cuanto al nivel en que pueden usarse; por ejemplo, un incremento factible máximo de la altura de las chimeneas, pero también existe una gran flexibilidad para usar el método en cualquier nivel fraccionario de su límite tecnológico.

La tabla 3.13 muestra la cantidad de emisión (en millones de libras anuales) que se puede eliminar de cada tipo de horno mediante el empleo del método de abatimiento al máximo límite tecnológico. Para fines de análisis se supone que cada método se puede usar a un nivel menor para lograr cualquier fracción de reducción de las tasas de emisión que se presentan en esta tabla. Más aún, las fracciones pueden ser diferentes para los altos hornos y los hornos Siemens-Martin, y el uso simultáneo de otro método no afecta de manera significativa la reducción de emisiones que alcanza cada uno de ellos.

Después de obtener estos datos, quedó claro que ningún método por sí solo podía lograr las reducciones requeridas. Por otro lado, la combinación de los tres métodos a toda su capacidad —lo que sería demasiado caro si se quiere que los productos tengan precios competitivos— genera un resultado mucho más elevado de lo que se pide. Por todo esto, la conclusión de los ingenieros fue que debían usar alguna combinación de métodos, tal vez con capacidades fraccionarias, basada en sus costos relativos. Aún más, debido a las diferencias entre los altos hornos y los hornos Siemens-Martin, es probable que la combinación sea diferente para cada tipo de horno.

Se llevó a cabo un análisis para estimar el costo total anual de cada método de abatimiento. El costo anual de un método incluye el aumento de los gastos de operación y mantenimiento al igual que la reducción de los ingresos debida a cualquier pérdida de eficiencia en el proceso de producción que pueda generar el uso del método. El otro costo importante es el *costo fijo inicial* (el capital

TABLA 3.13 Reducción de la tasa de emisión (en millones de libras por año) con el uso máximo factible del método de abatimiento de Nori & Leets Co.

	Chimeneas más altas		Filtros		Mejores combustibles	
Contaminante	Altos hornos	Hornos Siemens-Martin	Altos hornos	Hornos Siemens-Martin	Altos hornos	Hornos Siemens-Martin
Partículas	12	9	25	20	17	13
Óxidos de azufre	35	42	18	31	56	49
Hidrocarbonos	37	53	28	24	29	20

[6] Un tiempo después de realizado el estudio, este método específico de reducción de contaminación se convirtió en materia de controversia. Como su efecto es reducir la contaminación a nivel del suelo esparciendo las emisiones a una distancia mayor, los grupos ecologistas sostienen que este procedimiento crea más lluvia ácida pues mantiene los óxidos de azufre más tiempo en el aire. En consecuencia, la U.S. Environmental Protection Agency adoptó nuevas reglas en 1985 para eliminar los incentivos sobre el uso de chimeneas más altas.

TABLA 3.14 Costo total anual por el uso máximo factible del método de abatimiento de Nori & Leets Co. (millones de dólares)

Método de abatimiento	Altos hornos	Hornos de corazón abierto
Chimeneas más altas	8	10
Filtros	7	6
Mejores combustibles	11	9

inicial) que se requiere para instalar el método. Para hacer que este costo único fuera conmensurable con los costos anuales, se usó el valor del dinero en el tiempo para calcular el gasto anual (sobre el tiempo esperado de vida del método) que sería equivalente a este costo fijo inicial.

El análisis permitió estimar los costos anuales totales (en millones de dólares), que se presentan en la tabla 3.14, en que se incurre al usar los métodos a toda su capacidad de abatimiento. También se determinó que el costo de un método que se utiliza a un nivel menor es esencialmente proporcional a la capacidad fraccional de la capacidad de abatimiento que se logra, aspecto que se presenta en la tabla 3.13. Entonces, para cualquier fracción que se logre, el costo total anual sería en esencia la fracción de la cantidad correspondiente de la tabla 3.14.

En esta etapa, todo está listo para desarrollar el marco general del plan de la compañía para disminuir la contaminación. Este plan especifica qué tipo de métodos de reducción deberán emplearse y a qué fracciones de su capacidad para: 1) los altos hornos y 2) los hornos Siemens-Martin. Debido a la naturaleza combinatoria del problema de encontrar un plan que satisfaga los requisitos con el menor costo posible, se formó un equipo de investigación de operaciones para resolverlo. El equipo decidió enfocar el problema desde un punto de vista de programación lineal, y formuló el modelo que se resume a continuación.

Formulación como un problema de programación lineal. Este problema tiene seis variables de decisión, x_j, $j = 1, 2, \ldots, 6$, que representan el uso de cada uno de los tres métodos de reducción en cada tipo de horno, expresado como una *fracción de la capacidad de reducción* (de manera que x_j no exceda de 1). En la tabla 3.15 se muestra el orden asignado a estas variables. Tomándose en cuenta que el objetivo es minimizar el costo total sin violar los requerimientos de reducción de la emisión, los datos de las tablas 3.12, 3.13 y 3.14 condujeron al siguiente modelo:

$$\text{Minimizar} \quad Z = 8x_1 + 10x_2 + 7x_3 + 6x_4 + 11x_5 + 9x_6,$$

sujeta a las siguientes restricciones:

1. Reducción de emisión:

$$\begin{aligned} 12x_1 + 9x_2 + 25x_3 + 20x_4 + 17x_5 + 13x_6 &\geq 60 \\ 35x_1 + 42x_2 + 18x_3 + 31x_4 + 56x_5 + 49x_6 &\geq 150 \\ 37x_1 + 53x_2 + 28x_3 + 24x_4 + 29x_5 + 20x_6 &\geq 125 \end{aligned}$$

2. Tecnológicas:

$$x_j \leq 1, \quad \text{para } j = 1, 2, \ldots, 6$$

TABLA 3.15 Variables de decisión (fracción del uso máximo factible del método de abatimiento) de Nori & Leets Co.

Método de abatimiento	Altos hornos	Hornos de corazón abierto
Chimeneas más altas	x_1	x_2
Filtros	x_3	x_4
Mejores combustibles	x_5	x_6

■ **TABLA 3.16** Datos de productos de la Save-It Co.

Grado	Especificación	Amalgamado, costo ($) por libra	Precio de venta ($) por libra
A	Material 1: no más de 30% del total Material 2: no menos de 40% del total Material 3: no más de 50% del total Material 4: exactamente 20% del total	3.00	8.50
B	Material 1: no más de 50% del total Material 2: no menos de 10% del total Material 4: exactamente 10% del total	2.50	7.00
C	Material 1: no más de 70% del total	2.00	5.50

3. No negatividad:

$$x_j \geq 0, \quad \text{para } j = 1, 2, \ldots, 6.$$

El equipo de IO usó este modelo[7] para encontrar el plan de costo mínimo

$$(x_1, x_2, x_3, x_4, x_5, x_6) = (1, 0.623, 0.343, 1, 0.048, 1),$$

con $Z = 32.16$ (costo total anual de 32.16 millones). Después realizó un análisis de sensibilidad para explorar el efecto de hacer los ajustes posibles en los estándares del aire que se presentan en la tabla 3.12, y para verificar el efecto de inexactitudes en los datos de costo dados en la tabla 3.14. (Esta historia continuará en el caso de estudio 6.1 al final del capítulo 6.) Después, se hizo una planeación detallada y la administración la aprobó. Muy poco tiempo después el programa se puso en práctica y los habitantes de Steeltown respiraron aire —más limpio— con alivio.

Reciclado de desechos sólidos

La SAVE-IT COMPANY opera un centro de reciclado que recoge cuatro tipos de material de desecho sólido y los trata para amalgamarlos en un producto que pueda lanzarse al mercado. El tratamiento y el amalgamado son dos procesos diferentes. Se pueden obtener tres grados diferentes de este producto (vea la primera columna de la tabla 3.16), según la mezcla de materiales que se use. Aunque existe alguna flexibilidad para esta mezcla en cada grado, los estándares de calidad especifican una cantidad mínima y una máxima de la proporción de los materiales permitidos en ese grado. (La proporción es el peso del material expresado como un porcentaje del peso total del producto de ese grado.) Para los dos grados más altos se especifica un porcentaje fijo de uno de los materiales. Estas especificaciones se presentan en la tabla 3.16 junto con el costo de amalgamado y el precio de venta de cada grado.

El centro de reciclado recoge los materiales de desecho sólido de ciertas fuentes habituales por lo que casi siempre puede mantener una tasa de producción estable para tratarlos. En la tabla 3.17 se muestran las cantidades disponibles para la recolección y tratamiento semanal, al igual que el costo del proceso de cada tipo de material.

La Save-It Co. es propiedad de Green Earth, una organización dedicada a asuntos ecológicos, por lo que las ganancias se usan para apoyar las actividades de Green Earth. Esta organización ha logrado contribuciones y apoyos por la cantidad de 30 000 dólares semanales, que deben usarse sólo para cubrir el costo del tratamiento completo de los desechos sólidos. El consejo directivo de Green Earth ha girado instrucciones a la administración de Save-It para que divida este dinero entre los materiales, de manera tal que se recolecte y se trate *al menos la mitad* de la cantidad disponible de cada tipo de material. Estas restricciones adicionales se enumeran en la tabla 3.17.

[7] Una formulación equivalente puede expresar cada variable de decisión en unidades naturales para su método de abatimiento; por ejemplo, x_1 y x_2 podrían representar el número de *pies* que aumentan las alturas de las chimeneas.

TABLA 3.17 Datos de los materiales de desechos sólidos de Save-It Co.

Material	Libras por semana disponible	Costo del tratamiento ($) por libra	Restricciones adicionales
1	3 000	3.00	1. De cada material deben recolectarse y tratarse al menos la mitad de las libras disponibles por semana.
2	2 000	6.00	2. Deben usarse $30 000 semanales para tratar estos materiales.
3	4 000	4.00	
4	1 000	5.00	

Con las restricciones especificadas en las tablas 3.16 y 3.17, la administración desea determinar la *cantidad* que debe producir de cada grado *y* la *mezcla* exacta de materiales que usará para cada uno, de manera que se maximice la ganancia semanal neta —ingresos totales por ventas *menos* costo total del amalgamado—, independiente del costo del tratamiento fijo de 30 000 dólares por semana que será cubierto por donaciones.

Formulación como un problema de programación lineal. Antes de intentar construir un modelo de programación lineal debe tenerse mucho cuidado en la definición apropiada de las variables de decisión. Si bien muchas veces esta definición es obvia, otras es la parte medular de la formulación. Después de identificar con claridad cuál es la información que sirve y la forma más conveniente de manejarla mediante las variables de decisión, se pueden establecer la función objetivo y las restricciones sobre los valores de estas variables de decisión.

En este problema específico, las decisiones que deben tomarse están bien definidas, pero vale la pena pensar un poco en la manera de manejar la información a través de ellas. (Se recomienda intentar hacerlo y ver si primero se obtiene el siguiente conjunto *inapropiado* de variables de decisión.)

Como un conjunto de decisiones se refiere a la *cantidad* de cada grado de producto que se debe fabricar, parecería natural definir un conjunto de variables de decisión acorde. Siguiendo tentativamente esta línea de pensamiento se define

$$y_i = \text{número de libras del producto de grado } i \text{ producidas por semana} \quad (i = A, B, C).$$

El otro conjunto de decisiones es la *mezcla* de materiales de cada grado de producto. Esta mezcla se identifica por la proporción de cada material en el producto, lo que sugiere definir el otro conjunto de variables de decisión como

$$z_{ij} = \text{proporción del material } j \text{ en el producto de grado } i \quad (i = A, B, C;\ j = 1, 2, 3, 4).$$

Sin embargo, en la tabla 3.17 se proporcionan los costos del tratamiento y la disponibilidad de los materiales por *cantidad* (libras) y no en *proporciones*, y es esta información en *cantidades* la que se necesita registrar en algunas de las restricciones. Para el material j ($j = 1, 2, 3, 4$),

$$\text{Cantidad en libras de material } j \text{ usado por semana} = z_{Aj}y_A + z_{Bj}y_B + z_{Cj}y_C.$$

Por ejemplo, como la tabla 3.17 indica que se dispone de 3 000 libras del material 1 por semana, una restricción del modelo sería

$$z_{Aj}y_A + z_{Bj}y_B + z_{Cj}y_C \leq 3\,000.$$

Desafortunadamente, ésta no es una restricción legítima de programación lineal. La expresión del lado izquierdo *no es* una función lineal porque incluye la multiplicación de variables. Por lo tanto, no se puede construir un modelo de programación lineal con estas variables de decisión.

Por suerte, existe otra manera de definirlas que se ajusta al formato de programación lineal. (¿Es posible identificar alguna manera de hacerlo?) Este objetivo se logra con sólo sustituir

cada *producto* de las variables de decisión anteriores por una sola variable. En otras palabras se define

$$x_{ij} = z_{ij}y_i \qquad (\text{para } i = A, B, C; j = 1, 2, 3, 4)$$
$$= \text{número total de libras del material } j \text{ asignadas al producto grado } i \text{ por semana,}$$

y después se definen x_{ij} como las variables de decisión. Al combinar las x_{ij} en diferentes formas se llega a las siguientes cantidades necesarias en el modelo (para $i = A, B, C; j = 1, 2, 3, 4$).

$$x_{i1} + x_{i2} + x_{i3} + x_{i4} = \text{número de libras del producto grado } i \text{ fabricado por semana.}$$
$$x_{Aj} + x_{Bj} + x_{Cj} = \text{número de libras de material } j \text{ usado por semana.}$$
$$\frac{x_{ij}}{x_{i1} + x_{i2} + x_{i3} + x_{i4}} = \text{proporción del material } j \text{ en el producto de grado } i.$$

El hecho de que esta última expresión sea una función *no lineal* no causa complicaciones. Por ejemplo, considere la primera especificación para el producto grado A en la tabla 3.16 (la proporción de material 1 no debe exceder de 30%). Esta limitación conduce a la restricción no lineal

$$\frac{x_{A1}}{x_{A1} + x_{A2} + x_{A3} + x_{A4}} \leq 0.3.$$

Sin embargo, al multiplicar ambos lados de esta desigualdad por el denominador se llega a la restricción *equivalente*

$$x_{A1} \leq 0.3(x_{A1} + x_{A2} + x_{A3} + x_{A4}),$$

de manera que

$$0.7x_{A1} - 0.3x_{A2} - 0.3x_{A3} - 0.3x_{A4} \leq 0,$$

es una restricción legítima de programación lineal.

Con este ajuste a las tres cantidades dadas se determinan directamente todas las restricciones funcionales del modelo. La función objetivo se basa en la meta de la administración de maximizar la ganancia semanal total —ingresos totales por ventas *menos* costo total del amalgamado— que se obtiene por los tres grados de productos. En consecuencia, la ganancia por libra de cada grado de producto se obtiene mediante la resta del costo del amalgamado que se presenta en la tercera columna de la tabla 3.16 del precio de venta que aparece en la cuarta columna. Estas *diferencias* proporcionan los coeficientes de la función objetivo.

Por lo tanto, el modelo completo de programación lineal es

$$\text{Maximizar} \quad Z = 5.5(x_{A1} + x_{A2} + x_{A3} + x_{A4}) + 4.5(x_{B1} + x_{B2} + x_{B3} + x_{B4})$$
$$+ 3.5(x_{C1} + x_{C2} + x_{C3} + x_{C4}),$$

sujeta a las siguientes restricciones:

1. Especificaciones de mezcla (segunda columna de la tabla 3.16):

$$x_{A1} \leq 0.3(x_{A1} + x_{A2} + x_{A3} + x_{A4}) \quad (\text{grado } A, \text{ material } 1)$$
$$x_{A2} \geq 0.4(x_{A1} + x_{A2} + x_{A3} + x_{A4}) \quad (\text{grado } A, \text{ material } 2)$$
$$x_{A3} \leq 0.5(x_{A1} + x_{A2} + x_{A3} + x_{A4}) \quad (\text{grado } A, \text{ material } 3)$$
$$x_{A4} = 0.2(x_{A1} + x_{A2} + x_{A3} + x_{A4}) \quad (\text{grado } A, \text{ material } 4)$$
$$x_{B1} \leq 0.5(x_{B1} + x_{B2} + x_{B3} + x_{B4}) \quad (\text{grado } B, \text{ material } 1)$$
$$x_{B2} \geq 0.1(x_{B1} + x_{B2} + x_{B3} + x_{B4}) \quad (\text{grado } B, \text{ material } 2)$$
$$x_{B4} = 0.1(x_{B1} + x_{B2} + x_{B3} + x_{B4}) \quad (\text{grado } B, \text{ material } 4)$$
$$x_{C1} \leq 0.7(x_{C1} + x_{C2} + x_{C3} + x_{C4}) \quad (\text{grado } C, \text{ material } 1).$$

2. Disponibilidad de materiales (segunda columna de la tabla 3.17):

$$x_{A1} + x_{B1} + x_{C1} \le 3\,000 \quad \text{(material 1)}$$
$$x_{A2} + x_{B2} + x_{C2} \le 2\,000 \quad \text{(material 2)}$$
$$x_{A3} + x_{B3} + x_{C3} \le 4\,000 \quad \text{(material 3)}$$
$$x_{A4} + x_{B4} + x_{C4} \le 1\,000 \quad \text{(material 4).}$$

3. Restricciones sobre las cantidades tratadas (lado derecho de la tabla 3.17):

$$x_{A1} + x_{B1} + x_{C1} \ge 1\,500 \quad \text{(material 1)}$$
$$x_{A2} + x_{B2} + x_{C2} \ge 1\,000 \quad \text{(material 2)}$$
$$x_{A3} + x_{B3} + x_{C3} \ge 2\,000 \quad \text{(material 3)}$$
$$x_{A4} + x_{B4} + x_{C4} \ge 500 \quad \text{(material 4).}$$

4. Restricción sobre el costo del tratamiento (lado derecho de la tabla 3.17):

$$3(x_{A1} + x_{B1} + x_{C1}) + 6(x_{A2} + x_{B2} + x_{C2}) + 4(x_{A3} + x_{B3} + x_{C3}) + 5(x_{A4} + x_{B4} + x_{C4}) = 30\,000.$$

5. Restricciones de no negatividad:

$$x_{A1} \ge 0, \quad x_{A2} \ge 0, \quad \ldots, \quad x_{C4} \ge 0.$$

Esta formulación completa el modelo, excepto que las restricciones de las especificaciones de la mezcla necesitan reescribirse en la forma adecuada para un modelo de programación lineal con todas las variables en el lado izquierdo y combinar los términos:

Especificaciones de mezcla:

$$0.7x_{A1} - 0.3x_{A2} - 0.3x_{A3} - 0.3x_{A4} \le 0 \quad \text{(grado } A\text{, material 1)}$$
$$-0.4x_{A1} + 0.6x_{A2} - 0.4x_{A3} - 0.4x_{A4} \ge 0 \quad \text{(grado } A\text{, material 2)}$$
$$-0.5x_{A1} - 0.5x_{A2} + 0.5x_{A3} - 0.5x_{A4} \le 0 \quad \text{(grado } A\text{, material 3)}$$
$$-0.2x_{A1} - 0.2x_{A2} - 0.2x_{A3} + 0.8x_{A4} = 0 \quad \text{(grado } A\text{, material 4)}$$
$$0.5x_{B1} - 0.5x_{B2} - 0.5x_{B3} - 0.5x_{B4} \le 0 \quad \text{(grado } B\text{, material 1)}$$
$$-0.1x_{B1} + 0.9x_{B2} - 0.1x_{B3} - 0.1x_{B4} \ge 0 \quad \text{(grado } B\text{, material 2)}$$
$$-0.1x_{B1} - 0.1x_{B2} - 0.1x_{B3} + 0.9x_{B4} = 0 \quad \text{(grado } B\text{, material 4)}$$
$$0.3x_{C1} - 0.7x_{C2} - 0.7x_{C3} - 0.7x_{C4} \le 0 \quad \text{(grado } C\text{, material 1).}$$

La tabla 3.18 muestra una solución óptima para este modelo, y después estos valores de x_{ij} se usan para calcular otras cantidades de interés dadas en la misma tabla. El valor óptimo de la función objetivo que se obtiene es $Z = 35\,109.65$ (o sea, una ganancia semanal total de \$35 109.65).

■ **TABLA 3.18 Solución óptima del problema de Save-It Co.**

	Libras usadas por semana				
	Material				
Grado	**1**	**2**	**3**	**4**	**Número de libras producidas por semana**
A	412.3 (19.2%)	859.6 (40%)	447.4 (20.8%)	429.8 (20%)	2 149
B	2 587.7 (50%)	517.5 (10%)	1 552.6 (30%)	517.5 (10%)	5 175
C	0	0	0	0	0
Total	3 000	1 377	2 000	947	

Recuadro de aplicación

El control de costos es esencial para la supervivencia en la industria aérea. Por lo tanto, la alta administración de **United Airlines** inició un estudio de investigación de operaciones para mejorar la utilización de personal en las oficinas de reservaciones de la aerolínea y en los aeropuertos al ajustar de manera más cercana los programas de trabajo con las necesidades del cliente. El número de empleados necesario en cada ubicación para proporcionar el nivel de servicio que se requiere varía en gran medida durante las 24 horas del día y puede fluctuar de manera considerable cada media hora.

El diseño de programas de trabajo para todos los empleados en una ubicación dada para satisfacer estas necesidades de servicio de manera más eficiente es una pesadilla de consideraciones combinatorias. Una vez que el empleado llega, estará ahí de manera continua durante todo el turno (entre dos y 10 horas, dependiendo del empleado), *con excepción* de la hora de la comida y los descansos cada dos horas. Dado el número *mínimo* de empleados que trabajan que se requieren para *cada* intervalo de media hora en un día de 24 horas (este mínimo cambia día con día en una semana de siete días), ¿*cuántos* empleados de *cada longitud de turno* deben comenzar a trabajar en *qué momento de inicio cada día* de 24 horas en una semana de siete días? Por fortuna, la programación lineal trata con estas pesadillas combinatorias. El modelo de programación lineal para algunas de las ubicaciones programadas implicaron ¡más de 20 000 decisiones!

Esta aplicación de la programación lineal recibió el crédito por *ahorrarle a United Airlines más de* **6 millones de dólares** *anuales* sólo por salarios directos y costos de prestaciones. Entre otros beneficios obtenidos están la mejora del servicio al cliente y la reducción de las cargas de trabajo para el equipo de apoyo.

Fuente: T. J. Holloran y J. E. Bryne, "United Airlines Station Manpower Planning System", en *Interfaces*, **16**(1): 39-50, enero-febrero, 1986. (En nuestro sitio web www.mhhe.com/hillier se proporciona un vínculo con este artículo.)

El problema de la Save-It Co. es ejemplo de un **problema de mezclas**. El objetivo de un problema de este tipo es encontrar la mejor mezcla de ingredientes de los productos finales para cumplir con ciertas especificaciones. Algunas de las primeras aplicaciones de programación lineal se hicieron para la *mezcla de gasolina*, en donde los ingredientes del petróleo se mezclaban para obtener varios grados de gasolina. El reconocido estudio de IO en Texaco que se presenta al final de la sección 2.5 trata sobre mezcla de gasolina. Otros problemas de mezclas incluyen productos finales como acero, fertilizantes y alimento para animales.

Programación de personal

UNION AIRWAYS va a agregar vuelos desde y hacia su aeropuerto base, por lo cual necesita contratar más agentes de servicio a clientes. Sin embargo, no está claro cuántos más debe contratar. La administración reconoce la necesidad de controlar el costo y al mismo tiempo proporcionar de manera permanente un nivel satisfactorio de servicio. Por todo esto, un equipo de IO estudia la forma de programar a los agentes para proporcionar un servicio satisfactorio con el menor costo en personal.

Con base en la nueva programación de vuelos, se ha realizado un análisis del número *mínimo* de agentes de servicio a clientes que deben encontrarse de guardia en diferentes momentos del día para proporcionar un nivel satisfactorio de servicio. La columna de la derecha de la tabla 3.19 muestra el número de agentes necesario para los periodos dados en la primera columna. Los otros datos de la tabla reflejan uno de los acuerdos del contrato colectivo vigente entre la compañía y el sindicato que representa a los agentes de servicio a clientes. El acuerdo es que cada agente trabaje un turno de 8 horas 5 días a la semana, y los turnos autorizados son

Turno 1: 6:00 a.m. a 2:00 p.m.
Turno 2: 8:00 a.m. a 4:00 p.m.
Turno 3: 12:00 a.m. (mediodía) a 8:00 p.m.
Turno 4: 4:00 p.m. a 12 p.m. (medianoche)
Turno 5: 10:00 p.m. a 6:00 a.m.

Las marcas en el cuerpo principal de la tabla 3.19 muestran las horas cubiertas por los turnos respectivos. Como algunos turnos son menos deseables que otros, los salarios que se especifican en el contrato difieren de uno a otro. En el último renglón se muestra la compensación diaria —con las prestaciones— por cada agente para cada turno. El problema consiste en determinar cuántos agentes deben asignarse a los turnos respectivos cada día para minimizar el costo *total* de personal debido a los agentes, de acuerdo con este último renglón, al mismo tiempo que se cumplen (o se sobrepasan) las necesidades de servicio dados en la columna de la extrema derecha.

■ **TABLA 3.19 Datos del problema de programación de personal en Union Airways**

	Periodos cubiertos					
	Turno					
Periodo	**1**	**2**	**3**	**4**	**5**	**Número mínimo necesario de agentes**
6:00 a.m. a 8:00 a.m.	✔					48
8:00 a.m. a 10:00 a.m.	✔	✔				79
10:00 a.m. a 12 a.m.	✔	✔				65
12 a.m. a 2:00 p.m.	✔	✔	✔			87
2:00 p.m. a 4:00 p.m. .		✔	✔			64
4:00 p.m. a 6:00 p.m.			✔	✔		73
6:00 p.m. a 8:00 p.m.			✔	✔		82
8:00 p.m. a 10:00 p.m..				✔		43
10:00 p.m. a 12:00 p.m.				✔	✔	52
12:00 p.m. a 6:00 a.m.					✔	15
Costo diario por agente	$170	$160	$175	$180	$195	

Formulación como un problema de programación lineal. Los problemas de programación lineal siempre implican encontrar la mejor *mezcla de los niveles de actividad*. La clave para formular este problema en particular es reconocer la naturaleza de las actividades.

Las *actividades* corresponden a los turnos, donde el nivel de cada actividad es el número de agentes asignados a ese turno. Por lo tanto, este problema trata de encontrar la mejor *mezcla de tamaños de turnos*. Como las variables de decisión siempre son los niveles de actividades, las cinco variables de decisión en este caso son

x_j = número de agentes asignados al turno j, para $j = 1, 2, 3, 4, 5$.

La restricción principal sobre los valores de estas variables de decisión es que el número de agentes que trabaja en cada periodo debe satisfacer el requerimiento mínimo que se presenta en la columna de la derecha de la tabla 3.19. Por ejemplo, de las 2:00 p.m. a las 4:00 p.m., el número total de agentes asignados a los turnos que cubren este periodo (turnos 2 y 3) debe ser al menos 64, de manera que

$$x_2 + x_3 \geq 64$$

es la restricción funcional para este periodo.

Como el objetivo es minimizar el costo total de los agentes asignados a los cinco turnos, los coeficientes de la función objetivo se presentan en el último renglón de la tabla 3.19.

Por lo tanto, el modelo completo de programación lineal es

$$\text{Minimizar} \quad Z = 170x_1 + 160x_2 + 175x_3 + 180x_4 + 195x_5,$$

sujeta a

$$
\begin{array}{lll}
x_1 & \geq 48 & \text{(6-8 a.m.)} \\
x_1 + x_2 & \geq 79 & \text{(8-10 a.m.)} \\
x_1 + x_2 & \geq 65 & \text{(10-12 a.m.)} \\
x_1 + x_2 + x_3 & \geq 87 & \text{(12 a.m.-2 p.m.)} \\
x_2 + x_3 & \geq 64 & \text{(2-4 p.m.)} \\
x_3 + x_4 & \geq 73 & \text{(4-6 p.m.)} \\
x_3 + x_4 & \geq 82 & \text{(6-8 p.m.)} \\
x_4 & \geq 43 & \text{(8-10 p.m.)} \\
x_4 + x_5 & \geq 52 & \text{(10-12 p.m.)} \\
x_5 & \geq 15 & \text{(12 p.m.-6 a.m.)}
\end{array}
$$

y

$$x_j \geq 0, \quad \text{para } j = 1, 2, 3, 4, 5.$$

Si se observa con cuidado se ve que la tercera restricción, $x_1 + x_2 \geq 65$, en realidad no es necesaria porque la segunda, $x_1 + x_2 \geq 79$, asegura que $x_1 + x_2$ será mayor que 65. Así, $x_1 + x_2 \geq 65$ es una restricción *redundante* que se puede eliminar. De manera similar, la sexta restricción $x_3 + x_4 \geq 73$, también es *redundante* porque la séptima es $x_3 + x_4 \geq 82$. (En realidad, tres de las restricciones de no negatividad, $x_1 \geq 0$, $x_4 \geq 0$, $x_5 \geq 0$, también son restricciones redundantes debido a la primera, la octava y la décima: $x_1 \geq 48$, $x_4 \geq 43$ y $x_5 \geq 15$. Sin embargo, no se facilitan los cálculos si se eliminan estas restricciones de no negatividad.)

La solución óptima para este modelo es $(x_1, x_2, x_3, x_4, x_5) = (48, 31, 39, 43, 15)$. Esto conduce a $Z = 30\,610$, es decir, un costo total diario de personal de $30 610.

Este problema es un ejemplo donde en realidad no se satisface el supuesto de divisibilidad de un modelo de programación lineal. El número de agentes asignados a cada turno necesita ser entero. Si se habla de manera estricta, el modelo debe tener una restricción adicional para cada una de las variables de decisión que especifique que sus valores deben ser enteros. Al agregar estas restricciones, el modelo de programación lineal se convierte en un modelo de programación entera (tema del capítulo 11).

Aun sin estas restricciones, resulta que la solución óptima dada tiene valores enteros, por lo cual su no inclusión no generó problemas. (La forma de las restricciones funcionales permitió que este resultado fuera probable.) Si alguna de las variables hubiera sido no entera, el enfoque más sencillo habría sido *redondearla* a un valor entero. (El redondeo es factible en este ejemplo porque todas las restricciones funcionales son de la forma ≥ con coeficientes no negativos.) Cuando se redondea no necesariamente se obtiene una solución óptima para el modelo de programación lineal entera, pero el error que se introduce con algunos números grandes es despreciable en la mayoría de las situaciones prácticas. De manera alternativa, se pueden usar las técnicas de programación entera descritas en el capítulo 11 para obtener una solución óptima exacta con valores enteros.

El segundo Recuadro de aplicación en esta sección describe la manera como United Airlines aplicó programación lineal para desarrollar un sistema de programación de personal de tamaño mucho mayor que el ejemplo.

Distribución de bienes a través de una red

El problema. La DISTRIBUTION UNLIMITED CO. fabricará el mismo nuevo producto en dos plantas distintas y después tendrá que enviarlo a dos almacenes de distribución, donde cualquiera de las dos fábricas puede abastecer a cualquiera de los dos almacenes. La red de distribución disponible para el envío de este producto se muestra en la figura 3.13, donde Fl y F2 son las dos fábricas, Al y A2 son los dos almacenes y CD es el centro de distribución. Las cantidades que deben enviarse desde Fl y F2 se muestran a la izquierda, y las cantidades que deben recibirse en Al y A2 se presentan a la derecha. Cada flecha representa un canal factible de envío. Fl puede enviar directamente a Al y tiene tres rutas posibles (F1 → CD → A2, F1 → F2 → CD → A2 y F1 → A1 → A2) para mandar bienes a A2. La fábrica F2 tiene sólo una ruta a A2 (F2 → CD → A2) y una a A1 (F2 → CD → A2 → A1). El costo por unidad enviada a través de cada canal se muestra al lado de la flecha. También, junto a F1 → F2 y CD → A2 se muestran las cantidades máximas que se pueden enviar por estos canales. Los otros canales tienen suficiente capacidad para manejar todo lo que las fábricas pueden enviar.

La decisión que debe tomarse se refiere a qué cantidades enviar a través de cada canal de distribución. El objetivo es minimizar el costo total de envío.

Formulación como un problema de programación lineal. Con siete canales de envío se necesitan siete variables de decisión ($x_{F1\text{-}F2}, x_{F1\text{-}CD}, x_{F1\text{-}A1}, x_{F2\text{-}CD}, x_{CD\text{-}A2}, x_{A1\text{-}A2}, x_{A2\text{-}A1}$) para representar las cantidades enviadas a través de los canales respectivos.

Existen varias restricciones sobre los valores de estas variables. Además de las restricciones usuales de no negatividad, se tienen dos *restricciones de cota superior*, $x_{F1\text{-}F2} \leq 10$ y $x_{CD\text{-}A2} \leq 80$, impuestas por la capacidad limitada de envío de los dos canales, F1 → F2 y CD → A2. Todas las demás restricciones surgen de las cinco *restricciones de flujo neto*, una para cada localidad. Estas restricciones tienen la siguiente forma.

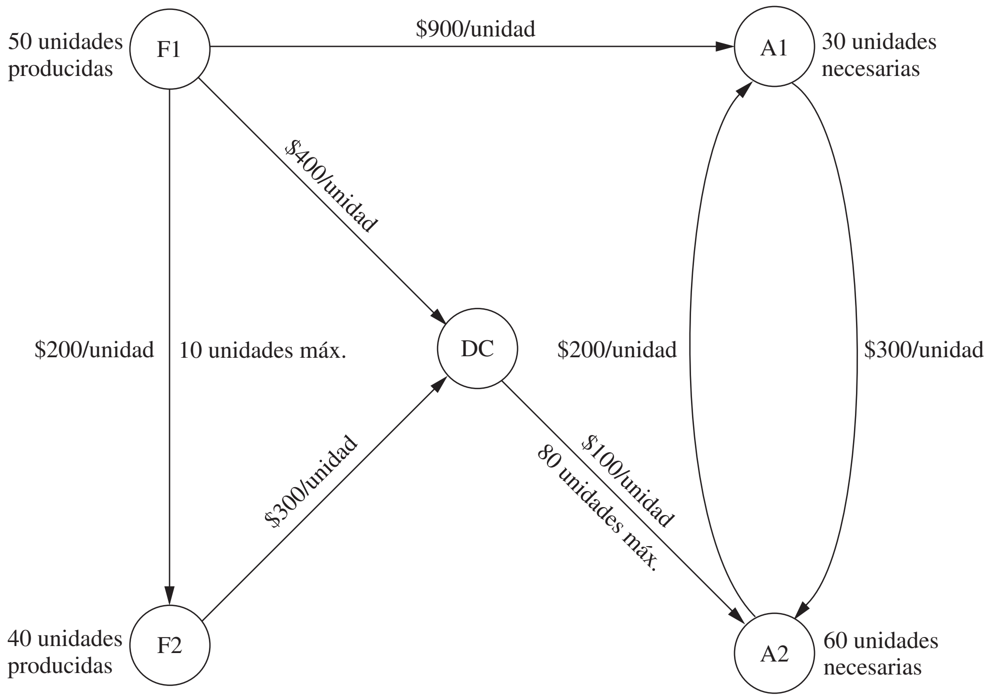


FIGURA 3.13
Red de distribución de Distribution Unlimited Co.

Restricción de flujo neto para cada localidad:

Cantidad enviada – cantidad recibida = cantidad requerida.

Como se indica en la figura 3.13, estas cantidades requeridas son 50 para F1, 40 para F2, –30 para A1 y –60 para A2.

¿Cuál es la cantidad requerida para CD? Todas las unidades producidas en las fábricas se necesitan en algún momento en los almacenes, de manera que las unidades enviadas de las fábricas a los centros de distribución deben mandarse a los almacenes. Por lo tanto, la cantidad total enviada de los centros de distribución a los almacenes debe ser *igual* a la cantidad total enviada desde las fábricas al centro de distribución. En otras palabras, la *diferencia* de estas dos cantidades enviadas —la cantidad requerida para la restricción de flujo neto— debe ser *cero*.

Como el objetivo es minimizar el costo total de envío, los coeficientes de la función objetivo son directamente los costos unitarios de envío que se muestran en la figura 3.13. Por lo tanto, si se usan unidades monetarias en cientos de dólares en esta función objetivo, el modelo completo de programación lineal es

$$\text{Minimizar} \quad Z = 2x_{\text{F1-F2}} + 4x_{\text{F1-CD}} + 9x_{\text{F1-A1}} + 3x_{\text{F2-CD}} + x_{\text{CD-A2}} + 3x_{\text{A1-A2}} + 2x_{\text{A2-A1}},$$

sujeta a las siguientes restricciones:

1. Restricciones de flujo neto:

$$\begin{aligned}
x_{\text{F1-F2}} + x_{\text{F1-CD}} + x_{\text{F1-A1}} &= 50 \text{ (fábrica 1)}\\
-x_{\text{F1-F2}} + x_{\text{F2-CD}} &= 40 \text{ (fábrica 2)}\\
-x_{\text{F1-CD}} - x_{\text{F2-CD}} + x_{\text{CD-A2}} &= 0 \text{ (centro de distribución)}\\
-x_{\text{F1-A1}} + x_{\text{A1-A2}} - x_{\text{A2-A1}} &= -30 \text{ (almacén 1)}\\
-x_{\text{CD-A2}} - x_{\text{A1-A2}} + x_{\text{A2-A1}} &= -60 \text{ (almacén 2)}
\end{aligned}$$

2. Restricciones de límite superior

$$x_{\text{F1-F2}} \leq 10, \quad x_{\text{CD-A2}} \leq 80$$

3. Restricciones de no negatividad:

$$x_{\text{F1-F2}} \geq 0, \quad x_{\text{F1-CD}} \geq 0, \quad x_{\text{F1-A1}} \geq 0, \quad x_{\text{F2-CD}} \geq 0, \quad x_{\text{CD-A2}} \geq 0,$$
$$x_{\text{A1-A2}} \geq 0, \quad x_{\text{A2-A1}} \geq 0.$$

Este problema se verá de nuevo en la sección 9.6, que está dedicada a problemas de programación lineal de este tipo, llamados *problemas de flujo de costo mínimo*. En la sección 9.7 se obtendrá su solución óptima:

$$x_{\text{F1-F2}} = 0, \quad x_{\text{F1-CD}} = 40, \quad x_{\text{F1-A1}} = 10, \quad x_{\text{F2-CD}} = 40, \quad x_{\text{CD-A2}} = 80,$$
$$x_{\text{A1-A2}} = 0, \quad x_{\text{A2-A1}} = 20.$$

El costo total de envío resultante es de $49 000.

3.5 FORMULACIÓN Y SOLUCIÓN DE MODELOS DE PROGRAMACIÓN LINEAL EN UNA HOJA DE CÁLCULO

Los paquetes de hojas de cálculo, como Excel, son una herramienta conocida para analizar y resolver problemas pequeños de programación lineal. Es sencillo introducir en una hoja de cálculo las características principales de un modelo de programación lineal, entre ellas, todos sus parámetros. Sin embargo, este software puede hacer mucho más que sólo desplegar datos. Si se incluye cierta información adicional, la hoja de cálculo se puede usar para analizar con rapidez soluciones potenciales. Por ejemplo, se puede verificar una solución potencial para ver si es factible y qué valor de Z (ganancia o costo) adquiere. Una gran parte del poder de la hoja de cálculo estriba en su capacidad para apreciar de inmediato los resultados de los cambios en la solución.

Además, el Excel Solver puede aplicar el método símplex para encontrar una solución óptima para el modelo.

Para ilustrar este proceso se considerará de nuevo el ejemplo de la Wyndor Glass de la sección 3.1.

Formulación del modelo en una hoja de cálculo

En la figura 3.14 se despliega el problema de Wyndor, sólo que se transfirieron los datos de la tabla 3.1 a una hoja de cálculo. (Las columnas E y F han sido reservadas para introducir otros datos que se describirán más adelante.) En lo sucesivo se hará referencia a las celdas que contienen los datos como **celdas de datos**. Estas celdas están sombreadas para distinguirlas de las otras que integran la hoja de cálculo.[8]

FIGURA 3.14 Hoja de cálculo inicial del problema de la Wyndor después de transferir los datos de la tabla 3.1 a las celdas de datos.

	A	B	C	D	E	F	G
1	Wyndor Glass Co. Product-Mix Problem						
2							
3			Doors	Windows			
4		Profit Per Batch	$3 000	$5 000			
5							Hours
6			Hours Used Per Batch Produced				Available
7		Plant 1	1	0			4
8		Plant 2	0	2			12
9		Plant 3	3	2			18

[8] Los bordes y el sombreado de celdas se pueden modificar mediante los botones de bordes y sombreado que están en la barra de herramientas de Formato o al seleccionar "Celdas" del menú "Formato" para después elegir la opción de "Bordes" o "Tramas".

Recuadro de aplicación

Welch's, Inc. es el procesador de uvas Concord y Niágara más grande del mundo, con ventas que sobrepasan 550 millones de dólares al año. Tales productos como la mermelada y el jugo de uva Welch han sido disfrutados por generaciones de consumidores estadounidenses.

Cada septiembre, los agricultores comienzan a entregar sus cosechas a las plantas de procesamiento que después exprimen las uvas crudas para obtener jugo. Debe pasar tiempo antes de que el jugo esté listo para convertirlo en jalea, mermelada, jugos y concentrados.

La decisión de cómo usar la cosecha de uva es una tarea compleja dada la demanda cambiante y la calidad y cantidad inciertas de la cosecha. Las decisiones típicas incluyen cuáles embarques deben usarse para elaborar los grupos de productos más importantes, la transferencia de jugo de uva entre las plantas y el modo de transporte para esta transferencia.

Debido a que Welch's no tenía un sistema formal para optimizar el movimiento de materia prima y los embarques usados para la producción, un equipo de IO desarrolló un modelo de programación lineal. Éste fue un modelo grande con 8 000 variables de decisión que se enfocaba en el nivel de detalle de los componentes. Las pruebas a pequeña escala probaron que el modelo funcionaba.

Para que el modelo fuera más útil, el equipo lo modificó agregando la demanda por grupo de productos en vez de por componentes. Esto redujo su tamaño a 324 variables de decisión y 361 restricciones funcionales. *Después el modelo se incorporó a una hoja de cálculo*.

La compañía ha utilizado de manera continua la versión actualizada de este *modelo de hoja de cálculo* cada mes desde 1994 para proporcionar a la alta administración información sobre el plan logístico generado por el Solver. Los *ahorros* por usar y optimizar este modelo *fueron de aproximadamente* **150 000 dólares** *sólo en el primer año*. Una ventaja importante de incorporar el modelo de programación lineal en una hoja de cálculo ha sido la facilidad de explicar el modelo a los administradores con diferentes niveles de comprensión matemática. Esto condujo a una valoración amplia del enfoque de investigación de operaciones tanto por esta aplicación como por otras.

Fuente: E.W. Schuster y S. J. Allen, "Raw Material Management at Welch's, Inc.", en *Interfaces*, **28**(5): 13-24, septiembre-octubre, 1998. (En nuestro sitio web www.mhhe.com/hillier se presenta un vínculo con este artículo.)

Después se puede observar que se ha facilitado su interpretación mediante el empleo de nombres de rango. Un **nombre de rango** es el título descriptivo que se da a un conjunto de celdas que de inmediato identifica lo que éstas contienen. Así, las celdas de datos en el problema de Wyndor han recibido los nombres de rango GananciaUnitaria (C4:D4), HorasUsadasPorLoteProducido (C7:D9), y HorasDisponibles (G7:G9). Observe que no se permiten los espacios en los nombres de rango y que cada nueva palabra comienza con una letra mayúscula. Aunque es opcional, el rango de celdas a las que se les da un nombre de rango puede especificarse entre paréntesis después del nombre. Por ejemplo, el rango C7:D9 es la forma corta en Excel para referirse al *rango desde* C7 *hasta* D9, es decir, todo el bloque de celdas en las columnas C o D y en los renglones 7, 8 o 9. Para introducir un nombre de rango, primero se debe seleccionar el rango de celdas, después se elige Nombre\Definir del menú Insertar y se digita un nombre de rango, o se da clic en el cuadro de nombre a la izquierda de la barra de fórmulas arriba de la hoja de cálculo y se digita un nombre.

Antes de iniciar el proceso de utilización de una hoja de cálculo para formular un modelo de programación lineal para el problema es necesario contestar tres preguntas.

1. ¿Qué *decisiones* deben tomarse? En este problema, las decisiones necesarias son las *tasas de producción* —número de lotes producidos a la semana— de los dos nuevos productos.
2. ¿Cuáles son las *restricciones* de estas decisiones? En este caso, son que el número de horas semanales de producción de los dos productos en las plantas respectivas no puede superar el número de horas disponibles.
3. ¿Cuál es la *medida global de desempeño* de estas decisiones? La medida de desempeño global de Wyndor es la *ganancia total* que se obtiene cada semana por los dos productos, por lo cual el *objetivo* es *maximizar* dicha cantidad.

En la figura 3.15 se muestra la forma como estas respuestas se pueden incorporar a una hoja de cálculo. Con base en la primera respuesta, las *tasas de producción* de los dos productos se colocan en las celdas C12 y D12 para ubicarlas en las columnas de estos productos justo debajo de las celdas de datos. Como aún no se sabe cuáles deben ser las tasas de producción, en este punto se introducen sólo como ceros. (En realidad, se puede introducir cualquier solución de prueba, aunque los valores *negativos* se deben excluir puesto que son imposibles.) Estos números cambiarán a medida que se busca la mejor mezcla de tasas de producción. Por ello, las celdas que contienen las decisiones

FIGURA 3.15
Hoja de cálculo completa del problema de la Wyndor con una solución de prueba inicial (ambas tasas de producción iguales a cero) introducidas en las celdas cambiantes (C12 y D12).

	A	B	C	D	E	F	G
1	Wyndor Glass Co. Product-Mix Problem						
2							
3			Doors	Windows			
4		Profit Per Batch	$3 000	$5 000			
5					Hours		Hours
6			Hours Used Per Batch Produced		Used		Available
7		Plant 1	1	0	0	<=	4
8		Plant 2	0	2	0	<=	12
9		Plant 3	3	2	0	<=	18
10							
11			Doors	Windows			Total Profit
12		Batches Produced	0	0			$0

que se deben tomar se llaman **celdas cambiantes** (o *celdas ajustables*). Para resaltar las celdas cambiantes, éstas se muestran sombreadas y con un borde. (En los archivos para hoja de cálculo del OR Courseware, las celdas cambiantes aparecen sombreadas en amarillo brillante.) Las celdas cambiantes han recibido el nombre de rango LotesProducidos (C12:D12).

Con base en la respuesta 2, el número total de horas de producción utilizadas cada semana por los dos productos en las plantas respectivas se introduce en las celdas E7, E8 y E9, justo a la derecha de las celdas de datos correspondientes. Las ecuaciones de Excel para estas tres celdas son

E7 = C7*C12 + D7*D12

E8 = C8*C12 + D8*D12

E9 = C9*C12 + D9*D12

donde cada asterisco denota multiplicación. Como cada una de estas celdas proporciona una salida que depende de las celdas cambiantes (C12 y D12) son llamadas **celdas de salida**.

Observe que cada una de las ecuaciones de las celdas de salida implica la suma de dos productos. Existe una función en Excel, llamada SUMAPRODUCTO, que suma el producto de los términos individuales que contienen dos diferentes rangos de celdas cuando los dos rangos tienen el mismo número de renglones y columnas. Cada producto que será sumado es el producto de un término del primer rango y el término de la ubicación correspondiente del segundo rango. Por ejemplo, considere los dos rangos, C7:D7 y C12:D12, de manera que cada rango tiene un renglón y dos columnas. En este caso, SUMAPRODUCTO (C7:D7, C12:D12) toma cada término individual del rango C7:D7, los multiplica por el término correspondiente del rango C12:D12 y después suma estos productos individuales, tal como se muestra en la ecuación anterior. Si se utiliza el nombre de rango LotesProducidos (C12:D12), la fórmula se convierte en SUMAPRODUCTO (C7:D7, LotesProducidos). Aunque es opcional en ecuaciones tan pequeñas, esta función es útil cuando se trata de capturar ecuaciones de programación lineal más largas.

Después se introducen los signos $\leq$ en las celdas F7, F8 y F9 para indicar que cada valor total a su izquierda no puede exceder al número correspondiente en la columna G. La hoja de cálculo permitirá introducir soluciones de prueba que violan los signos $\leq$. Sin embargo, estos signos sirven como recordatorio de que tales soluciones de prueba deben ser rechazadas si no se hacen cambios en los números de la columna G.

Por último, como la respuesta a la tercera pregunta es que la medida global de desempeño es igual a la ganancia total obtenida por los dos productos, dicha ganancia (por semana) se introduce en la celda G12. En forma parecida a los números de la columna E, ésta se obtiene mediante la suma de productos,

G12 = SUMAPRODUCTO (C4:D4, C12:D12)

Si se utilizan los nombres de rango GananciaTotal (G12), GananciaPorLote (C4:D4) y LotesProducidos (C12:D12), esta ecuación se convierte en

GananciaTotal = SUMAPRODUCTO (GananciaPorLote, LotesProducidos)

	A	B	C	D	E	F	G
1	Wyndor Glass Co. Product-Mix Problem						
2							
3			Doors	Windows			
4		Profit Per Batch	$3 000	$5 000			
5					Hours		Hours
6			Hours Used Per Batch Produced		Used		Available
7		Plant 1	1	0	0	<=	4
8		Plant 2	0	2	0	<=	12
9		Plant 3	3	2	0	<=	18
10							
11			Doors	Windows			Total Profit
12		Batches Produced	0	0			$0

Range Name	Cells
BatchesProduced	C12:D12
HoursAvailable	G7:G9
HoursUsed	E7:E9
HoursUsedPerBatchProduced	C7:D9
ProfitPerBatch	C4:D4
TotalProfit	G12

	E
5	Hours
6	Used
7	=SUMPRODUCT(C7:D7,BatchesProduced)
8	=SUMPRODUCT(C8:D8,BatchesProduced)
9	=SUMPRODUCT(C9:D9,BatchesProduced)

	G
11	Total Profit
12	=SUMPRODUCT(ProfitPerBatch,BatchesProduced)

FIGURA 3.16
Modelo en hoja de cálculo del problema de la Wyndor que incluye las fórmulas de la celda objetivo GananciaTotal (G12) y las otras celdas de salida en la columna E, donde el objetivo es maximizar la celda objetivo.

Éste es un buen ejemplo del beneficio que se obtiene cuando se utilizan los nombres de rango para que las ecuaciones resultantes sean más fáciles de interpretar. En lugar de tener que consultar la hoja de cálculo para ver lo que hay en las celdas G12, C4:D4 y C12:D12, los nombres de rango revelan de inmediato lo que hace la ecuación.

GananciaTotal (G12) es un tipo especial de celda de salida. Ésta es la celda particular que tiene el objetivo de alcanzar un valor tan grande como sea posible cuando se toman las decisiones relativas a las tasas de producción. Por lo tanto, GananciaTotal (G12) es referida como la **celda objetivo**. Dicha celda está sombreada con un color más oscuro que el de las celdas cambiantes y se distingue con un borde más grueso. (En los archivos para hoja de cálculo que contiene el OR Courseware, esta celda aparece en color naranja.)

La esquina inferior derecha de la figura 3.16 muestra las fórmulas que deben introducirse en la columna de HorasUsadas y en la celda de GananciaTotal. También se presenta un resumen de los nombres de rango (en orden alfabético) y las direcciones de celda correspondientes.

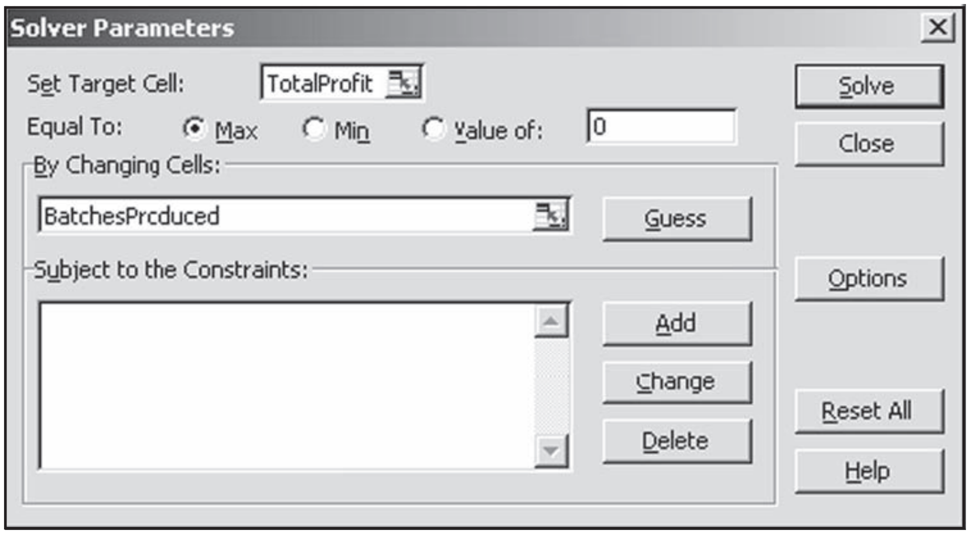
Lo anterior completa la formulación del modelo en una hoja de cálculo del problema de Wyndor.

Con esta formulación es sencillo analizar cualquier solución de prueba para las tasas de producción. Cada vez que se introducen tasas de producción en las celdas C12 y D12, Excel calcula de inmediato la cantidad total empleada de cada recurso y la ganancia total. Sin embargo, no es necesario utilizar la prueba y el error. A continuación se presenta cómo usar Excel para encontrar en forma rápida una solución óptima.

Uso de Excel Solver para resolver el modelo

Excel incluye una herramienta llamada Solver que aplica el método símplex para encontrar una solución óptima. (Una versión más poderosa de Solver, llamada *Premium Solver for Education*, está disponible en el OR Courseware.)

Para entrar a Solver por primera vez, es necesario ingresar al menú de complementos de Excel y agregar el Solver, y así podrá encontrarlo en la tabla de datos (Data tab) (para Excel 2007) o en el menú de herramientas (para versiones anteriores de Excel).



■ **FIGURA 3.17**
Cuadro de diálogo de Solver donde se especifica cuáles celdas de la figura 3.16 son las celdas objetivo y cuáles las que se modifican. También se indica que la celda objetivo será maximizada.

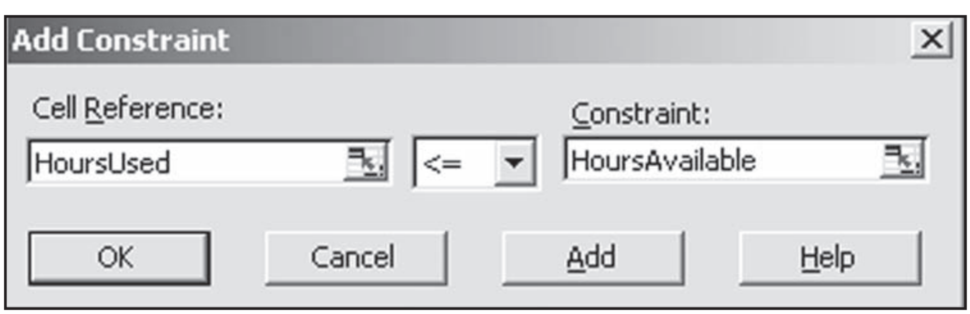
Para iniciar, en la figura 3.16 se ha introducido una solución arbitraria de prueba al colocar ceros en las celdas cambiantes. El Solver las modificará con los valores óptimos después de resolver el problema.

Este procedimiento comienza cuando se elige Solver del menú Herramientas. El cuadro de diálogo se muestra en la figura 3.17.

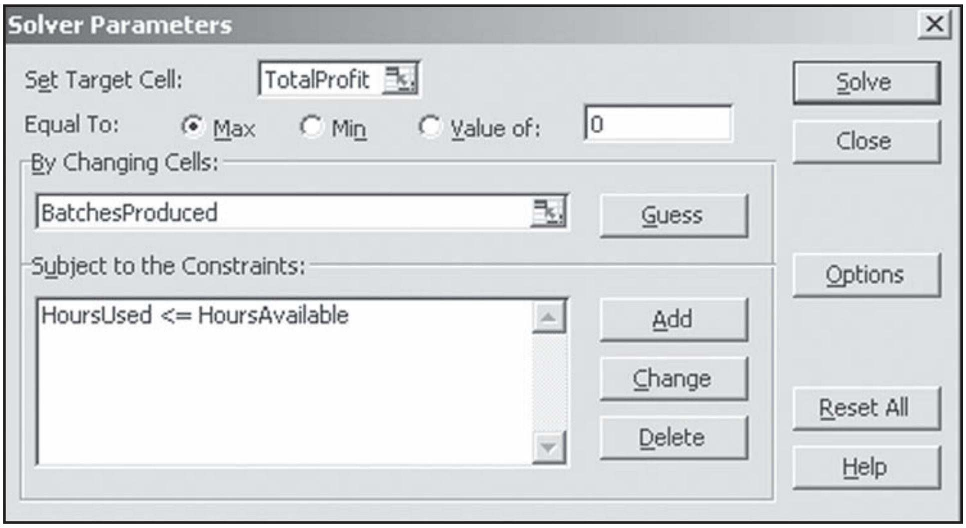
Antes de que Solver aplique el método símplex, necesita conocer con exactitud dónde se localizan los componentes del modelo en la hoja de cálculo. El cuadro de diálogo de Solver se usa para introducir esta información. El usuario tiene la opción de escribir los nombres de rango, las direcciones de las celdas o hacer clic sobre las celdas en la hoja de cálculo.[9] En la figura 3.17 se muestra el resultado de utilizar la primera opción, por lo cual se ha introducido GananciaTotal (en lugar de G12) para la celda objetivo y se ha escrito LotesProducidos (en lugar de rango C12:D12) para las celdas cuyo contenido cambiará. Como el objetivo es maximizar la meta de la celda objetivo, también se debe seleccionar Max.

Después deben especificarse las celdas que contienen las restricciones funcionales. Este objetivo se logra con un clic en el botón "agregar" en el cuadro de diálogo de Solver, con lo que aparece el cuadro de diálogo de "agregar restricción", que se muestra en la figura 3.18. Los signos ≤ en las celdas F7, F8 y F9 de la figura 3.16 son un recordatorio de que las celdas de HorasUsadas (E7:E9)

■ **FIGURA 3.18**
Cuadro de diálogo de Agregar restricción después de introducir el conjunto de restricciones, HorasUsadas (E7:E9) ≤ HorasDisponibles (G7:G9), las cuales especifican que se requiere que las celdas E7, E8 y E9 de la figura 3.16 sean menores o iguales que las celdas respectivas G7, G8 y G9.



[9] Si las celdas se seleccionan mediante un clic sobre ellas, aparecerán al principio en el cuadro de diálogo con sus direcciones y signos monetarios (por ejemplo, $C$9:$D$9), los cuales pueden pasarse por alto. Solver reemplazará tanto las direcciones de celda como los signos monetarios por el nombre de rango correspondiente (si es que existe algún nombre de rango definido para las celdas especificadas), pero sólo después de agregar una restricción o cerrar y reabrir el cuadro de diálogo de Solver.



■ **FIGURA 3.19**
Cuadro de diálogo de Solver después de especificar el modelo completo en términos de la hoja de cálculo.

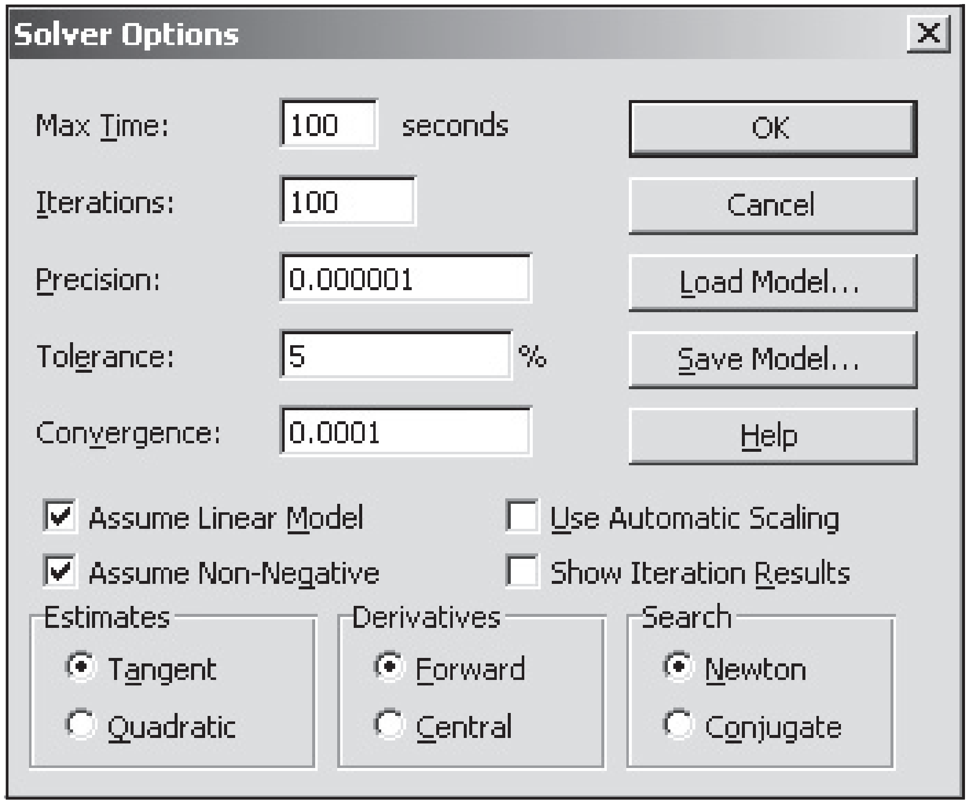
deben ser menores o iguales que las celdas correspondientes de HorasDisponibles (G7:G9). Estas restricciones se especifican en Solver al introducir HorasUsadas (o E7:E9) en el lado izquierdo del cuadro de diálogo de Agregar Restricción y HorasDisponibles (o G7:G9) en el lado derecho. Para ingresar el signo entre estos dos lados existe un menú para elegir entre los signos <= (menor o igual que), = o >= (mayor o igual que), en el que se ha elegido <= para estas restricciones. Esta selección es necesaria aun cuando antes se hayan escrito los signos ≤ en la columna F de la hoja de cálculo, porque Solver usa sólo las restricciones funcionales que se especifican en el cuadro de diálogo de Agregar restricción.

Si hubiera más restricciones funcionales que añadir se debería hacer clic en "agregar" para que aparezca de nuevo el cuadro de diálogo. Sin embargo, como no hay más en este ejemplo, el siguiente paso es hacer clic en "aceptar" para regresar al cuadro de diálogo de Solver.

Ahora el cuadro de diálogo resume el modelo completo (vea la figura 3.19) en términos de la hoja de cálculo de la figura 3.16. Sin embargo, antes de pedir a Solver que resuelva el modelo, se debe dar otro paso. Al elegir el botón "opciones" aparecerá el cuadro de diálogo que se muestra en la figura 3.20. Éste permite especificar cierto número de opciones sobre la manera en que se resolverá el modelo. Las más importantes son las opciones "adoptar modelo lineal" y "asumir no negativos". Asegúrese de elegir los cuadros de ambas opciones como se muestra en la figura. Esto indica a Solver que es un problema de programación *lineal* con restricciones de no negatividad para todas las variables de decisión, y que el método símplex se debe usar para resolver el problema. En cuanto al resto de las opciones, en general los valores preestablecidos se pueden aceptar para problemas pequeños. Con un clic en aceptar se regresa al cuadro de diálogo de Solver.

Ahora todo está listo para hacer clic en "resolver" en el cuadro de diálogo de Solver, con lo que se ejecutará el método símplex. Después de unos segundos (en el caso de un problema pequeño), Solver indicará los resultados. Por lo común señalará que encontró una solución óptima, como se especifica en el cuadro de diálogo de Resultados de Solver que se muestra en la figura 3.21. Si el modelo no tiene soluciones factibles o una solución óptima, este cuadro de diálogo lo indicará con un mensaje como "Solver no pudo encontrar una solución factible" o "El conjunto de valores en las celdas no converge". El cuadro de diálogo también presenta la opción de generar varios informes. Uno de ellos —el informe de sensibilidad— se analizará con detalle en las secciones 4.7 y 6.8.

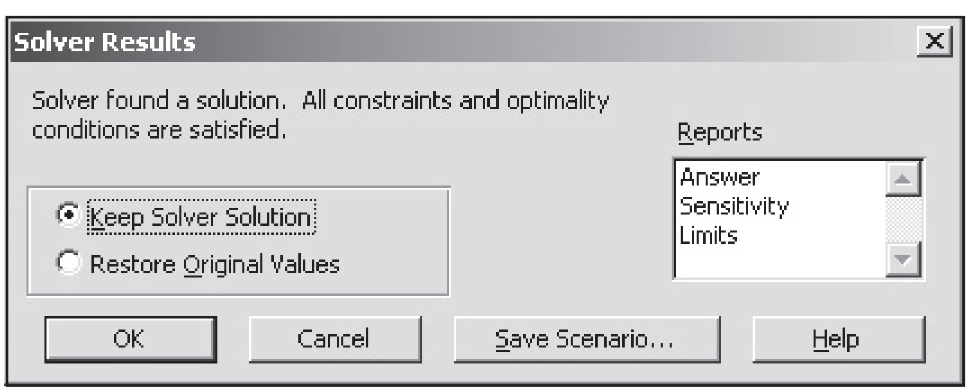
Después de resolver el modelo, el Solver sustituye el valor original de las variables de decisión en la hoja de cálculo por los valores óptimos, como se muestra en la figura 3.22. Así, la solución óptima es producir dos lotes de puertas y seis lotes de ventanas por semana, tal como se encontró mediante el método gráfico en la sección 3.1. La hoja de cálculo también indica el número correspondiente en la celda objetivo —una ganancia total de 36 000 dólares por semana—, así como los números en las celdas de salida HorasUsadas (E7:E9).



■ **FIGURA 3.20**
Cuadro de diálogo de Opciones de Solver después de elegir las opciones "Adoptar modelo lineal" y "Asumir no negativos" para indicar que se desea resolver un modelo de programación lineal que tiene restricciones de no negatividad.

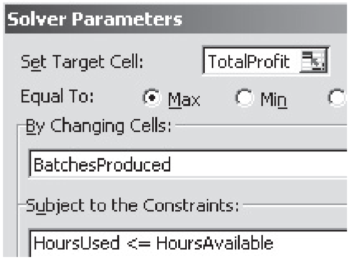
En este punto se podría verificar lo que pasa con la solución óptima si cualquiera de los números de las celdas de datos se cambiara por otros valores posibles. Este cambio es fácil de hacer porque el Solver guarda todas las direcciones de la celda objetivo, celdas cambiantes, restricciones, etc., cuando se guarda el archivo. Todo lo que se debe hacer es realizar los cambios que se desee en las celdas de datos y después dar clic de nuevo sobre "resolver" en el cuadro de diálogo del Solver. (Las secciones 4.7 y 6.8 se enfocan en este tipo de *análisis de sensibilidad*, incluso se aborda la manera de utilizar el Informe de Sensibilidad del Solver para realizar el análisis "qué pasaría si".)

Para ayudar a experimentar con este tipo de cambios, el OR Courseware incluye archivos de Excel para éste y otros capítulos que proporcionan una formulación completa y una solución de los ejemplos —el problema de Wyndor y los de la sección 3.4— en un formato de hoja de cálculo. Se recomienda "jugar" con estos ejemplos para ver qué pasa con los diferentes datos, distintas soluciones, etc. Estas hojas de cálculo también pueden resultar útiles como plantillas para resolver problemas de tarea.



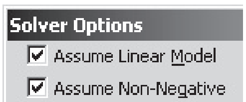
■ **FIGURA 3.21**
Cuadro de diálogo de Resultados de Solver que indica que se encontró una solución óptima.

	A	B	C	D	E	F	G
1		**Wyndor Glass Co. Product-Mix Problem**					
2							
3			Doors	Windows			
4		Profit Per Batch	$3 000	$5 000			
5					Hours		Hours
6			Hours Used Per Batch Produced		Used		Available
7		Plant 1	1	0	2	<=	4
8		Plant 2	0	2	12	<=	12
9		Plant 3	3	2	18	<=	18
10							
11			Doors	Windows			Total Profit
12		Batches Produced	2	6			$36 000



	E
5	Hours
6	Used
7	=SUMPRODUCT(C7:D7,BatchesProduced)
8	=SUMPRODUCT(C8:D8,BatchesProduced)
9	=SUMPRODUCT(C9:D9,BatchesProduced)

	G
11	Total Profit
12	=SUMPRODUCT(ProfitPerBatch,BatchesProduced)



Range Name	Cells
BatchesProduced	C12:D12
HoursAvailable	G7:G9
HoursUsed	E7:E9
HoursUsedPerBatchProduced	C7:D9
ProfitPerBatch	C4:D4
TotalProfit	G12

■ **FIGURA 3.22**
Hoja de cálculo obtenida después de resolver el problema de la Wyndor.

Además, se sugiere utilizar los archivos de Excel de este capítulo para revisar de manera cuidadosa las formulaciones en hoja de cálculo para algunos de los ejemplos de la sección 3.4. Este enfoque explica cómo formular modelos de programación lineal en hojas de cálculo cuando éstos son más grandes y complicados que el del problema de Wyndor.

En capítulos posteriores se verán otros ejemplos de cómo formular y resolver varios tipos de modelos de IO en una hoja de cálculo. Los capítulos complementarios del sitio web del libro también incluyen un capítulo completo (21) que está dedicado al modelado en hojas de cálculo. Ese capítulo describe en detalle tanto el proceso general como las directrices básicas para construir modelos en hoja de cálculo. También presenta algunas técnicas para depurar tales modelos.

3.6 CONSTRUCCIÓN DE MODELOS GRANDES DE PROGRAMACIÓN LINEAL

Los modelos de programación lineal son de muchos tamaños diferentes. Los ejemplos de las secciones 3.1 y 3.4 oscilan desde tres restricciones funcionales y dos variables de decisión —para la Wyndor y la terapia de radiación— hasta 17 restricciones funcionales y 12 variables de decisión —en el problema de Save-It Company—. El último caso puede parecer un modelo algo grande. Después de todo, lleva bastante tiempo elaborar uno de este tamaño. Sin embargo, los modelos de los recuadros de aplicación que se presentan en este capítulo son mucho más grandes. Por ejemplo, el modelo para la aplicación de United Airlines en la sección 3.4 tiene más de 20 000 variables de decisión.

Los modelos de ese tamaño no son del todo raros. Es común que, en la práctica, los modelos de programación lineal tengan cientos o miles de restricciones funcionales. En realidad, a veces tienen incluso millones de restricciones funcionales. Con frecuencia, el número de variables de

decisión es tan grande como el número de restricciones funcionales, y en ocasiones también llegan a ser millones.

La formulación de modelos tan grandes puede ser una tarea desalentadora. Aun un modelo de "tamaño medio" con mil restricciones y mil variables tiene más de un millón de parámetros, que incluyen el millón de coeficientes de estas restricciones. Sencillamente no es práctico elaborar la formulación algebraica ni introducir los parámetros en una hoja de cálculo para un modelo de este tipo.

Entonces, ¿cómo se construyen estos modelos grandes en la práctica? Se requiere el uso de un *lenguaje de modelado*.

Lenguajes de modelado

Un lenguaje de modelado de programación matemática es un software de diseño especial para formular de modo eficiente los modelos de programación lineal grandes, y otros relacionados. Incluso cuando se tienen miles de restricciones funcionales, éstas relativamente son de pocos tipos, y las del mismo tipo siguen el mismo patrón. De igual manera, las variables de decisión estarán dentro de unas cuantas categorías. Por ello, si se usan grandes bloques de datos en bases de datos, un lenguaje de modelado construirá todas las restricciones del mismo tipo a la vez, con fundamento en las variables de cada tipo. Este proceso se ilustrará en breve.

Además de formular con eficiencia los modelos grandes, un lenguaje de modelado facilita las tareas de administración del modelo, inclusive el acceso a los datos, su transformación en parámetros del modelo, la modificación del modelo cuando se desee y el análisis de las soluciones. También puede producir informes resumidos en el lenguaje de los tomadores de decisiones, al igual que documentar el contenido del modelo.

Se han desarrollado varios lenguajes de modelado excelentes en las últimas dos décadas, entre los que se destacan AMPL, MPL, GAMS y LINGO.

La versión para estudiantes de uno de ellos, **MPL** (siglas de mathematical programming language), está incluida en el sitio web del libro junto con un amplio material de ayuda. A medida que las versiones subsecuentes salgan al mercado en el futuro, se podrán bajar del sitio de Internet maximalsoftware.com. MPL es un producto de Maximal Software, Inc. Su nueva característica es el gran apoyo de Excel en MPL, el cual incluye tanto la importación como la exportación de conjuntos de datos desde y hacia Excel. También se proporciona un apoyo amplio para el lenguaje de macros de Excel VBA mediante el OptiMax 2000. (La versión para estudiantes de OptiMax 2000 se encuentra también en el sitio web del libro.) Este producto permite al usuario integrar totalmente los modelos de MPL en Excel y resolverlos con cualquiera de los poderosos solucionadores que soporta MPL, como **CPLEX** (descrito en la sección 4.8).

LINGO es un producto de LINDO Systems, Inc., que también comercializa un optimizador orientado por completo a hojas de cálculo llamado What's*Best!* que está diseñado para enfrentar grandes problemas industriales, así como una subrutina adicional llamada LINDO API. El software LINGO incluye una interfaz con el subprograma LINDO que ha sido una introducción popular a la programación lineal para muchas personas. La versión para estudiante de LINGO con la interfaz LINDO es parte del software incluido en el sitio web del libro. Todos los productos de Sistemas LINDO también se pueden bajar de www.lindo.com. Al igual que MPL, LINGO es un poderoso lenguaje de modelado de propósito general. Una característica notable de LINGO es su gran flexibilidad para tratar con una amplia variedad de problemas de IO además de los de programación lineal. Por ejemplo, cuando se abordan problemas complicados de programación no lineal, contiene un optimizador global que obtiene una solución globalmente óptima. (Comentarios adicionales a esto aparecen en la sección 12.10.) En esta nueva edición del libro, el LINGO más reciente también tiene un lenguaje de programación incluido de manera que es posible realizar tareas como resolver diferentes problemas de optimización como parte de una corrida, lo cual es particularmente útil cuando se hace análisis paramétrico,

El sitio web del libro incluye formulaciones de MPL, LINGO, LINDO y What's*Best!* para casi todos los ejemplos de este libro a los que se pueden aplicar estos lenguajes de modelado y optimizadores.

A continuación se presenta un ejemplo simplificado que ilustra cómo puede surgir un modelo de programación lineal muy grande.

Ejemplo de un problema con un modelo muy grande

La administración de WORLDWIDE CORPORATION desea resolver un *problema de mezcla de productos*, que es mucho más complejo que el problema de la Wyndor presentado en la sección 3.1. Esta corporación tiene 10 plantas en varias partes del mundo. Cada una elabora los mismos 10 productos y después los vende en su región. Se conoce la *demanda* (ventas potenciales) de cada producto en cada planta en cada uno de los próximos 10 meses. Aunque la cantidad de producto vendido en un mes dado no puede exceder la demanda, la cantidad producida puede ser mayor, y la cantidad en exceso se debería almacenar en inventario (con un costo unitario por mes) para su venta posterior. Cada unidad de cada producto ocupa el mismo espacio en almacén, y cada planta tiene un límite superior para el número total de unidades que se puede guardar (la *capacidad del inventario*).

Cada planta tiene los mismos 10 procesos de producción, se hará referencia a ellos como *máquinas*, cada uno de los cuales se puede usar para producir cualquiera de los 10 productos. Tanto el costo de producción por unidad como la tasa de producción de un producto —número de unidades producidas por día dedicado a ese producto— dependen de la combinación de plantas y máquinas involucradas (pero no del mes). El número de días hábiles (*días de producción disponibles*) varía un poco de un mes a otro.

Como algunas plantas y máquinas pueden producir un producto dado ya sea a menor costo o a una tasa más rápida que otras plantas y máquinas, en ocasiones vale la pena enviar algunas unidades del producto de una planta a otra para que esta última las venda. Existe cierto costo asociado con cada unidad enviada de cualquier producto de cada combinación de una planta que envía (*planta origen*) y una planta que recibe (*planta destino*), donde este costo unitario es el mismo para todos los productos.

La administración necesita determinar cuántas unidades de cada producto debe producir en cada máquina de cada planta cada mes, al igual que cuántas unidades de cada producto debe vender cada planta cada mes y cuántas unidades de cada producto debe enviar cada planta cada mes a cada una de las otras plantas. Si se toma en cuenta el precio en todo el mundo de cada producto, el objetivo es encontrar el plan factible que maximice la ganancia total: ingreso por ventas totales *menos* la suma de los costos totales de producción, inventario y envío.

Se debe considerar de nuevo que éste es un ejemplo simplificado en varias formas. Se ha supuesto que el número de plantas, máquinas, productos y meses es exactamente el mismo (10). En situaciones más reales es probable que el número de productos sea mucho más grande y que el horizonte de planeación sea mucho más grande que 10 meses; asimismo, el número de "máquinas" (tipos de procesos de producción) debería ser menor a 10. También se ha supuesto que todas las plantas tienen los mismos tipos de máquinas (procesos de producción) y que cada tipo de máquina puede fabricar todos los productos. En realidad, las plantas pueden tener algunas diferencias en términos de tipos de máquinas y de los productos que pueden manufacturar. El resultado neto es que el modelo correspondiente para algunas corporaciones puede ser más pequeño que el de este ejemplo, pero el modelo para otras puede ser mucho más grande.

Estructura del modelo resultante

Debido a los costos de inventario y a que las capacidades de almacenamiento son limitadas, es necesario mantener un registro de la cantidad de cada producto que se guarda en cada planta durante cada mes. En consecuencia, el modelo de programación lineal tiene cuatro tipos de variables de decisión: cantidades de producción, cantidades de inventario, cantidades de venta y cantidades enviadas. Con 10 plantas, 10 máquinas, 10 productos y 10 meses, esto da un total de 21 000 variables de decisión, como se describe a continuación.

Variables de decisión

10 000 variables de producción: una por cada combinación de planta, máquina, producto y mes
1 000 variables de inventario: una por cada combinación de planta, producto y mes
1 000 variables de ventas: una por cada combinación de planta, producto y mes
9 000 variables de envío: una por cada combinación de producto, mes, planta (planta origen, *fromplant*) y otra planta (la planta destino, *toplant*).

Cuando se multiplica cada variable de decisión por el costo unitario o ingreso unitario correspondiente y después se suma según cada tipo, se puede calcular la siguiente función objetivo:

Función objetivo

Maximizar ganancia = ingresos totales por ventas – costo total,

donde

Costo total = costo total de producción + costo total de inventario + costo total de envío.

Cuando se maximiza esta función objetivo, las 21 000 variables de decisión deben satisfacer las restricciones de no negatividad al igual que los cuatro tipos de restricciones funcionales: de capacidad de producción, de balanceo de plantas (restricciones de igualdad que proporcionan valores adecuados para las variables de inventario), de inventario máximo y de ventas máximas. Como se enumeran en seguida, existe un total de 3 100 restricciones funcionales, pero todas las de un mismo tipo siguen el mismo patrón.

Restricciones funcionales

1 000 restricciones de capacidad de producción (una por cada combinación de planta, máquina y mes):

Días de producción usados ≤ días de producción disponibles,

donde el lado izquierdo es la suma de 10 fracciones, una por cada producto, donde cada fracción es la cantidad de ese producto (una variable de decisión) *dividida* entre la tasa de producción del producto (una constante dada).

1 000 restricciones de balance de las plantas (una por cada combinación de planta, producto y mes):

Cantidad producida + inventario del mes pasado + cantidad recibida = ventas + inventario actual + cantidad enviada,

donde la *cantidad producida* es la suma de las variables de decisión que representan las cantidades de producción de las máquinas, la *cantidad recibida* es la suma de las variables de decisión que representan las cantidades enviadas desde otras plantas y la *cantidad enviada* es la suma de las variables de decisión correspondientes a las cantidades que se mandan a las otras plantas.

100 restricciones de inventario máximo (una por cada combinación de planta y mes):

Inventario total ≤ capacidad de inventario,

donde el lado izquierdo es la suma de las variables de decisión que representan las cantidades de inventario de los productos individuales.

1 000 restricciones de ventas (una por cada combinación de planta, producto y mes):

Ventas ≤ demanda.

Ahora se estudiará la manera en que el lenguaje de modelado MPL, un producto de Maximal Software, Inc., puede formular este enorme modelo en forma muy compacta.

Formulación del modelo en MPL

Quien construye el modelo comienza por asignarle un título y enumerar un *índice* por cada elemento del problema, como se ilustra en seguida.

```
TITLE
  Production_Planning;
INDEX
  product   := (A1, A2, A3, A4, A5, A6, A7, A8, A9, A10);
  month     := (Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct);
```

```
plant     :=(p1, p2, p3, p4, p5, p6, p7, p8, p9, p10);
fromplant := plant;
toplant   :=plant;
machine   :=(m1, m2, m3, m4, m5, m6, m7, m8, m9, m10);
```

Excepto por los meses, los elementos del lado derecho son etiquetas arbitrarias de los respectivos productos, plantas y máquinas, donde las mismas etiquetas se usan en los archivos de datos. Observe que se colocan dos puntos después del nombre de cada elemento y punto y coma al final de cada instrucción (pero una instrucción puede extenderse más de un renglón).

El trabajo pesado en cualquier modelo grande es la recolección y organización de los distintos tipos de datos en archivos de datos, los cuales pueden estar en formato denso o disperso. En el *formato denso*, el archivo contiene una entrada para cada posible combinación de los índices sobre los que corren los datos. Por ejemplo, suponga que el archivo de datos contiene las tasas de producción para elaborar los diferentes productos con las distintas máquinas (procesos de producción) en las diversas plantas. En formato denso, el archivo contiene una entrada para cada combinación de una planta, una máquina y un producto. Sin embargo, la entrada puede necesitar ser cero para la mayoría de las combinaciones porque esa planta en particular puede no tener esa máquina específica o, incluso si la tiene, esa máquina individual puede no ser capaz de producir ese producto particular en esa planta en específico. El porcentaje de las entradas en formato denso que son *distintas de cero* se conoce como densidad del conjunto de datos. En la práctica, es común que los conjuntos de datos tengan una *densidad* de datos por debajo de 5%, y a veces está por debajo de 1%. Los conjuntos de datos que tienen una densidad tan baja se dice que están dispersos. En tales situaciones, resulta más eficiente utilizar un archivo de datos en *formato disperso*. En este formato sólo se introducen al archivo de datos los valores distintos de cero (y una identificación de los valores índice a los que se refieren). Por lo general, los datos se leen en formato disperso desde un archivo de texto o de las bases de datos corporativas. La capacidad de manejar conjuntos de datos dispersos de manera eficiente es una de las claves para el éxito de la formulación y la resolución de modelos de optimización a gran escala. MPL puede trabajar con facilidad tanto con el formato denso como con el disperso.

En el ejemplo de la Worldwide Corp. se necesitan ocho archivos para precios de productos, demandas, costos de producción, tasas de producción, días de producción disponibles, costos de inventario, capacidades de inventario y costos de envío. Se supone que estos archivos de datos están disponibles en formato disperso. El siguiente paso es dar un nombre sugestivo corto a cada uno e identificar (entre paréntesis cuadrados) el índice o índices sobre los que corren los datos en los archivos como se muestra a continuación.

```
DATA
Price[product]      := SPARSEFILE("Price.dat");
Demand[plant,   product,    month] := SPARSEFILE("Demand.dat");
ProdCost[plant, machine, product] := SPARSEFILE("Produce.dat", 4);
ProdRate[plant, machine, product] := SPARSEFILE("Produce.dat", 5);
ProdDaysAvail[month]     := SPARSEFILE("ProdDays.dat");
InvtCost[plant, product] := SPARSEFILE("InvtCost.dat");
InvtCapacity[plant]      := SPARSEFILE("InvtCap.dat");
ShipCost[fromplant, toplant] := SPARSEFILE ("ShipCost.dat");
```

Para ilustrar lo que contienen estos archivos de datos, considere el que proporciona los costos y las tasas de producción. En seguida se presenta una muestra pequeña de las primeras entradas del archivo disperso produce.dat:

```
!
    ! Produce.dat - Production Cost and Rate
    !
    ! ProdCost[plant, machine, product]:
    ! ProdRate[plant, machine, product]:
    !
      p1, m11, A1, 73.30, 500,
      p1, m11, A2, 52.90, 450,
      p1, m12, A3, 65.40, 550,
      p1, m13, A3, 47.60, 350,
```

Después, el modelador da un nombre corto a cada tipo de variable de decisión. Después del nombre, entre paréntesis cuadrados, se escribe el (los) índice(s) sobre el (los) que corre el subíndice.

```
VARIABLES
  Produce[plant, machine, product, month]     -> Prod;
  Inventory[plant, product, month]            -> Invt;
  Sales[plant, product, month]                -> Sale;
  Ship[product, month, fromplant, toplant]
      WHERE (fromplant <> toplant);
```

En el caso de variables de decisión con nombres de más de cuatro letras, las flechas de la derecha señalan las abreviaturas de cuatro letras que se ajustan a las limitaciones de tamaño de varios solucionadores. La última línea indica que los subíndices de la planta origen y la planta destino no pueden tener el mismo valor.

Se debe dar un paso más antes de elaborar el modelo. A fin de que sea más sencillo leerlo es útil introducir primero *macros* que representan las sumas en la función objetivo.

```
MACROS
  Total Revenue := SUM(plant, product, month: Price*Sales);
  TotalProdCost := SUM(plant, machine, product, month:
                   ProdCost*Produce);
  TotalInvtCost := SUM(plant, product, month:
                   InvtCost*Inventory);
  TotalShipCost := SUM(product, month, fromplant, toplant:
                   ShipCost*Ship);
  TotalCost     := TotalProdCost + TotalInvtCost + TotalShipCost;
```

Los primeros cuatro macros usan la palabra reservada SUM de MPL para ejecutar las sumas involucradas. Después de cada palabra SUM (entre paréntesis) se coloca, primero, el (los) índice(s) sobre el (los) que corre la suma. Después (en seguida de dos puntos), el vector producto de un vector de datos (uno de los archivos de datos) multiplicado por un vector variable (uno de los cuatro tipos de variables de decisión).

Ahora, este modelo con 3 100 restricciones funcionales y 21 000 variables de decisión se puede escribir en la siguiente forma compacta.

```
MODEL
  MAX Profit = TotalRevenue - TotalCost;
SUBJECT TO
  ProdCapacity[plant, machine, month] -> PCap:
    SUM(product: Produce/ProdRate) <=  ProdDaysAvail;
  PlantBal[plant, product, month] -> PBal:
      SUM(machine: Produce) + Inventory [month - 1]
    + SUM(fromplant: Ship[fromplant, toplant: 5 plant])
  =
      Sales + Inventory
    + SUM(toplant: Ship[from plant: = plant, toplant]);

  MaxInventory [plant, month] -> MaxI:
    SUM(product: Inventory) <= InvtCapacity;
BOUNDS
    Sales <= Demand;
END
```

Por cada uno de los cuatro tipos de restricciones, el primer renglón da el nombre del tipo. Existe una restricción de este tipo para cada combinación de valores de los índices en los corchetes que siguen al nombre. A la derecha del paréntesis cuadrado, la flecha señala la abreviatura de cuatro letras que el solucionador puede usar. Abajo de este primer renglón se muestra la forma general de las restricciones de este tipo mediante el operador SUM.

Para cada restricción de capacidad de producción, los términos de la suma consisten en una variable de decisión (la cantidad de producción de ese producto en esa máquina en esa planta durante ese mes) dividida entre la tasa de producción correspondiente, que da el número de días

de producción empleados. Al sumar todos los productos se obtiene el número total de días de producción usados en esa máquina de esa planta durante ese mes, de modo que este número no debe exceder el número de días de producción disponibles.

El propósito de la restricción de balanceo de plantas de cada planta, producto y mes es proporcionar el valor correcto a la variable de inventario actual, dados los valores de todas las demás variables de decisión, incluso el nivel de inventario del mes anterior. En estas restricciones, cada operador SUM es sólo una suma de variables de decisión en vez de un producto vectorial. Ocurre lo mismo con el operador SUM en la restricción de inventario máximo. Por el contrario, el lado izquierdo de las restricciones de ventas máximas es una sola variable de decisión para cada una de las 1 000 combinaciones de planta, producto y mes. (La separación de estas restricciones de cota superior sobre las variables individuales de las restricciones funcionales normales tiene ventajas, debido a la eficiencia computacional que se logra cuando se emplea la *técnica de la cota superior* descrita en la sección 7.3.) No hay restricciones de cota inferior porque MPL supone de modo automático que las 21 000 variables de decisión tienen restricciones de no negatividad a menos que especifiquen cotas inferiores distintas de cero. En cada una de las 3 100 restricciones funcionales, observe que el lado izquierdo es una función lineal de las variables de decisión y que el lado derecho es una constante tomada del archivo de datos adecuado. Como la función objetivo también es una función lineal de las variables de decisión, se trata de un modelo legítimo de programación lineal.

Para resolver el modelo, MPL cuenta con varios **solucionadores** (paquetes de software para resolver modelos de programación lineal y modelos relacionados) que se pueden instalar en el MPL. Como se ve en la sección 4.8, **CPLEX** es un solucionador en especial poderoso. La versión de MPL en el OR Courseware ya tiene instalado la versión para estudiantes de CPLEX, que usa el método símplex para resolver estos modelos. Por lo tanto, para resolver un modelo formulado con MPL, todo lo que debe hacerse es elegir *Solve CPLEX* en el menú *Run* o seleccionar el botón *Run Solve* en la *barra de herramientas*. Después se puede ver el archivo con la solución en una ventana si se oprime el botón *View* en la parte inferior de la ventana *Status Window*.

Esta breve introducción a MPL ilustra la facilidad con que se pueden usar los lenguajes de modelado para formular modelos de programación lineal muy grandes de manera clara y concisa. Como ayuda cuando se utiliza MPL se incluye en el sitio web un tutorial, que revisa todos los detalles de construcción del modelo de las versiones más pequeñas de los ejemplos de planeación de la producción considerados aquí. En otro lugar del sitio web se puede ver cómo se formulan con MPL y se resuelven con CPLEX todos los ejemplos de programación lineal en este capítulo y los subsecuentes.

Lenguaje de modelado LINGO

LINGO es otro lenguaje de modelado conocido que se presenta en este libro. La compañía que lo produce, LINDO Systems, se dio a conocer primero por el optimizador **LINDO**, que es fácil de usar y es un subprograma del software LINGO. LINDO Systems también produce un solucionador en hoja de cálculo, ***What's Best!***, y una rutina solucionadora adicional, el **LINDO API**. La versión para el estudiante de LINGO se proporciona en el sitio web del libro. (Las versiones de prueba más recientes de todo lo anterior pueden bajarse de www.lindo.com.) Tanto LINDO como What's*Best!* comparten a LINDO API como el mecanismo de solución. El LINDO API tiene solucionadores basados en el método símplex y solucionadores de punto interior o barrera (como los que se estudian en las secciones 4.9 y 7.4), más un solucionador global para resolver modelos no lineales.

Como MPL, LINGO permite al modelador formular con eficiencia un modelo de programación lineal grande de manera clara y concisa que separa los datos de la formulación del modelo. Esta separación significa que a medida que ocurren cambios diarios en los datos que describen el problema que se debe resolver (los cambios pueden ser incluso minuto a minuto), el usuario sólo necesita cambiar los datos y no involucrarse con la formulación del modelo. Es posible desarrollar un modelo con un conjunto pequeño de datos y después, cuando se le agregan al modelo una cantidad de datos más grande, la formulación del modelo se ajusta de manera automática al nuevo conjunto de datos.

LINGO utiliza los *conjuntos* como sus conceptos fundamentales de construcción. Por ejemplo, en el problema de planeación de la producción de Worlwide Corp., los conjuntos de interés incluyen las colecciones de productos, plantas, máquinas y meses. Cada miembro de un conjunto puede tener uno o más *atributos* asociados, como el precio de un producto, la capacidad del in-

ventario de una planta, la tasa de producción de una máquina y el número de días de producción disponible en un mes. Algunos de estos atributos son datos de entrada, mientras que otros, como las cantidades de producción y embarque son variables de decisión del modelo. También es posible definir conjuntos derivados que se construyen a partir de combinaciones de otros conjuntos. Igual que en MPL, el operador SUM por lo común se usa para escribir la función objetivo y cada tipo de restricción de manera compacta.

Existe un manual impreso disponible de LINGO. Se puede acceder a él directamente en LINGO vía el comando Help, y puede explorarse de diferentes maneras.

Un suplemento de este capítulo en el sitio web del libro describe un poco más de LINGO e ilustra su uso con un par de ejemplos pequeños. Un segundo suplemento muestra cómo se puede usar LINGO para formular el modelo para el ejemplo de planeación de la producción de Worldwide Corp. Un tutorial de LINGO en el sitio web proporciona los detalles necesarios para realizar el modelado básico con este lenguaje. Las formulaciones y soluciones de LINGO para los distintos ejemplos tanto en este capítulo como en otros también se incluyen en el sitio web.

3.7 CONCLUSIONES

La programación lineal es una técnica poderosa para tratar problemas de asignación de recursos escasos entre actividades que compiten, al igual que otros problemas cuya formulación matemática es parecida. Se ha convertido en una herramienta estándar de gran importancia para muchas organizaciones industriales y de negocios. Aún más, casi cualquier organización social tiene el problema de asignar recursos en algún contexto y cada vez es mayor el reconocimiento de la aplicación tan amplia de esta técnica.

Sin embargo, no todos los problemas de asignación de recursos limitados se pueden formular de manera que se ajusten a un modelo de programación lineal, ni siquiera como una aproximación razonable. Cuando no se cumplen uno o más de los supuestos de programación lineal, tal vez sea posible aplicar otro tipo de modelos matemáticos, por ejemplo, los modelos de programación entera (capítulo 11) o de programación no lineal (capítulo 12).

REFERENCIAS SELECCIONADAS

1. Baker, K. R., *Optimization Modeling with Spreadsheets*, Duxbury, Pacific Grove, CA, 2006.
2. Hillier, F. S. y M. S. Hillier, *Introduction to Management Science: A Modeling and Case Studies Approach with Spreadsheets*, 2a. ed., McGraw-Hill/Irwin, Burr Ridge, IL, 2003, caps. 2, 4.
3. *LINGO User's Guide*, LINDO Systems, Inc., Chicago, IL, 2008.
4. Manual de *MPL Modeling System* (*Release* 4.2), Maximal Software, Inc., Arlington, VA, e-mail: info@maximalsoftware.com, 2008.
5. Schrage, L., *Optimization Modeling with LINGO*, LINDO Systems Press, Chicago, IL, 2003.
6. Williams, H. P., *Model Building in Mathematical Programming*, 4a. ed., Wiley, Nueva York, 1999.

Algunas aplicaciones de programación lineal ganadoras de premios

(En nuestro sitio web www.mhhe.com/hillier se proporciona un vínculo a todos estos artículos.)

A1. Ambs, K., S. Cwilich, M. Deng, D. J. Houck, D. F. Lynch y D. Yan, "Optimizing Restoration Capacity in the AT&T Network", en *Interfaces*, **30**(1): 26-44, enero-febrero de 2000.

A2. Caixeta-Filho, J. V., J. M. van Swaay-Neto y A. de P Wagemaker, "Optimization of the Production Planning and Trade of Lily Flowers at Jan de Wit Company", en *Interfaces*, **32**(1): 35-46, enero-febrero de 2002.

A3. Chalermkraivuth, K. C., S. Bollapragada, M. C. Clark, J. Deaton, L. Kiaer, J. P. Murdzek, W. Neeves, B. J. Scholz y D. Toledano, "GE Asset Managemet, Genworth Financial and GE Insurance Use a Sequential-Linear-Programming Algorithm to Optimize Portfolios", en *Interfaces*, **35**(5): 370-380, septiembre-octubre de 2005.

A4. Elimam, A. A., M. Girgis y S. Kotob, "A Solution to Post Crash Dedt Entanglements in Kuwait's al-Manakh Stock Market", en *Interfaces*, **27**(1): 89-106, enero-febrero de 1997.

A5. Epstein, R., R. Morales, J. Serón y A. Weintraub, "Use of OR Systems in the Chilean Forest Industries", en *Interfaces*, **29**(1): 7-29, enero-febrero de 1997.

A6. Geraghty, M. K. y E. Johnson, "Revenue Management Saves National Car Rental", en *Interfaces*, **27**l(1), 107-127, enero-febrero de 1997.

A7. Leachman, R. C., R. F. Benson, C. Liu y D. J, Raar, "IMPReSS: An Automed Production-Planning and Delivery-Quotation System at Harris Corporation—Semiconductor Sector", en Interfaces, **26**(1): 6-37, enero-febrero de 1996.
A8. Mukuch, W. M., J. L. Dodge, J. G. Ecker, D. C Granfors y G. J. Hahn, "Managing Consumer Credit Delinquency in the U. S. Economy: A Multi-Billion Dollar Management Science Application", en *Interfaces*, **22**(1): 90-109, enero-febrero de 1992.
A9. Murty, K. G., Y.-w. Wan, J. Liu, M. M. Tseng, E. Leung, K.-K. Lai y H. W. C. Chiu, "Hongkong International Terminals Gains Elastic Capacity Using a Data-Intensive Decision-Support System", en *Interfaces*, **35**(1): 61-75, enero-febrero de 2005.
A10. Yoshino, T., T. Sasaki y T. Hasegawa, "The Traffic-Control System on the Hanshin Expressway", en *Interfaces*, **25**(1): 94-108, enero-febrero de 1995.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN NUESTRO SITIO WEB (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos del capítulo 3

Ejemplo de demostración en OR Tutor:

Método gráfico

Procedimientos en IOR Tutorial:

Método gráfico interactivo
Método gráfico y análisis de sensibilidad

Complemento de Excel:

Premium Solver for Education

"Cap. 3. Introd. a archivos LP" para resolver los ejemplos:

Archivos Excel
Archivo LINGO/LINDO
Archivo MPL/CPLEX

Glosario del capítulo 3

Suplementos de este capítulo:

Lenguaje de modelado LINGO.
Más acerca de LINGO.

Vea el apéndice 1 de la documentación del software.

PROBLEMAS

Los símbolos a la izquierda de algunos problemas (o de sus incisos) significan lo siguiente:

D: El ejemplo de demostración indicado puede ser útil.
I: El uso del procedimiento correspondiente en IOR Tutorial puede resultar útil (la impresión registra su trabajo).
C: Utilice la computadora para resolver el problema con el método símplex. Las opciones de software disponibles incluyen Excel Solver o Premium Solver (sección 3.5). MPL/CPLEX (sección 3.6), LINGO (suplementos 1 y 2 de este capítulo en el sitio web del libro y apéndice 4.1) y LINDO (apéndice 4.1), pero siga las instrucciones de su profesor sobre cuál opción usar.

Un asterisco en el número del problema indica que al final del libro se da al menos una respuesta parcial.

3.1-1. Lea el artículo referenciado que describe todo el estudio de IO que se resume en el recuadro de aplicación que se presenta en la sección 3.1. Describa brevemente cómo se aplicó la programación lineal en este estudio. Después enumere los diferentes beneficios financieros y de otro tipo que resultaron de este estudio.

D **3.1-2.*** Para cada una de las siguientes restricciones, dibuje una gráfica individual para mostrar las soluciones no negativas que las satisfacen.

a) $x_1 + 3x_2 \leq 6$
b) $4x_1 + 3x_2 \leq 12$
c) $4x_1 + x_2 \leq 8$
d) Ahora combine estas restricciones en una sola gráfica para mostrar la región factible del conjunto completo de restricciones funcionales más las de no negatividad.

D **3.1-3.** Considere la siguiente función objetivo de un modelo de programación lineal:

Maximizar $Z = 2x_1 + 3x_2$

a) Dibuje en una gráfica las rectas correspondientes a la función objetivo de $Z = 6$, $Z = 12$ y $Z = 18$.
b) Encuentre la forma pendiente-ordenada al origen de la ecuación de estas tres rectas de la función objetivo. Compare las pendientes y las intersecciones con el eje x_2.

3.1-4. Considere la siguiente ecuación de una recta:

$60x_1 + 40x_2 = 600$

a) Encuentre la forma pendiente-ordenada al origen de esta ecuación.
b) Use esta forma para identificar la pendiente y la intersección de esta línea con el eje x_2.
c) Use la información del inciso *b*) para dibujar una gráfica de la recta.

D,I **3.1-5.*** Utilice el método gráfico para resolver el problema:

Maximizar $Z = 2x_1 + x_2$,

sujeta a

$x_2 \leq 10$

$$\begin{aligned} 2x_1 + 5x_2 &\leq 60 \\ x_1 + x_2 &\leq 18 \\ 3x_1 + x_2 &\leq 44 \end{aligned}$$

y

$x_1 \geq 0, \quad x_2 \geq 0.$

D,I **3.1-6.** Utilice el método gráfico para resolver el problema:

Maximizar $Z = 10x_1 + 20x_2$,

sujeta a

$$\begin{aligned} -x_1 + 2x_2 &\leq 15 \\ x_1 + x_2 &\leq 12 \\ 5x_1 + 3x_2 &\leq 45 \end{aligned}$$

y

$x_1 \geq 0, x_2 \geq 0.$

3.1-7. La empresa Whitt Window tiene sólo tres empleados que hacen dos tipos de ventanas a mano: con marco de madera y con marco de aluminio. La ganancia es de \$180 por cada ventana con marco de madera y de \$90 por cada una con marco de aluminio. Doug hace marcos de madera y puede terminar 6 al día. Linda hace 4 marcos de aluminio por día. Bob forma y corta el vidrio y puede hacer 48 pies cuadrados de vidrio por día. Cada ventana con marco de madera emplea 6 pies cuadrados de vidrio y cada una de aluminio, 8 pies cuadrados.

La compañía desea determinar cuántas ventanas de cada tipo debe producir al día para maximizar la ganancia total.

a) Describa la analogía entre este problema y el de Wyndor Glass Co. que se presentó en la sección 3.1. Después construya y llene una tabla como la 3.1 de este problema, e identifique las actividades y los recursos.
b) Formule un modelo de programación lineal para este problema.
D,I ***c***) Use el método gráfico para resolver el modelo.
I ***d***) Un nuevo competidor también produce ventanas con marco de madera. Esta circunstancia puede forzar a la compañía a reducir el precio y por ende la ganancia debida a este tipo de ventanas. ¿Cómo cambiaría la solución óptima (si cambia) si la ganancia por ventana de madera disminuyera de \$180 a \$120? ¿Y de \$180 a \$60? (Puede resultar útil emplear el procedimiento de análisis gráfico y análisis de sensibilidad del IOR Tutorial.)
I ***e***) Doug piensa disminuir sus horas de trabajo, lo cual reduciría el número de ventanas de madera que produce por día. ¿Cómo cambiaría la solución óptima si hace sólo 5 marcos diarios? (Puede resultar útil emplear el procedimiento de análisis gráfico y análisis de sensibilidad del IOR Tutorial.)

3.1-8. La compañía WorldLight produce dos dispositivos para lámparas (productos 1 y 2) que requieren partes de metal y componentes eléctricos. La administración desea determinar cuántas unidades de cada producto debe fabricar para maximizar la ganancia. Por cada unidad del producto 1 se requieren 1 unidad de partes de metal y 2 unidades de componentes eléctricos. Por cada unidad del producto 2 se necesitan 3 unidades de partes de metal y 2 unidades de componentes eléctricos. La compañía tiene 200 unidades de partes de metal y 300 de componentes eléctricos. Cada unidad del producto 1 da una ganancia de \$1 y cada unidad del producto 2, hasta 60 unidades, da una ganancia de \$2. Cualquier exceso de 60 unidades del producto 2 no genera ganancia, por lo que fabricar más de esa cantidad está fuera de consideración.

a) Formule un modelo de programación lineal.
D,I ***b***) Utilice el método gráfico para resolver este modelo. ¿Cuál es la ganancia total que resulta?

3.1-9. La compañía de seguros Primo está en proceso de introducir dos nuevas líneas de productos: seguro de riesgo especial e hipotecas. La ganancia esperada es de \$5 por el seguro de riesgo especial y de \$2 por unidad de hipoteca.

La administración desea establecer las cuotas de venta de las nuevas líneas para maximizar la ganancia total esperada. Los requerimientos de trabajo son los siguientes:

Departamento	Horas de trabajo por unidad: Riesgo especial	Horas de trabajo por unidad: Hipoteca	Horas de trabajo disponibles
Suscripciones	3	2	2 400
Administración	0	1	800
Reclamaciones	2	0	1 200

a) Formule un modelo de programación lineal.
D,I ***b***) Use el método gráfico para resolver el modelo.
c) Verifique el valor exacto de su solución óptima del inciso *b*) con la solución algebraica de las dos ecuaciones simultáneas relevantes.

3.1-10. Weenies and Buns es una planta procesadora de alimentos que fabrica hot dogs y pan para hot dogs. Muelen su propia harina a una tasa máxima de 200 libras por semana. Cada pan requiere 0.1 libras. Tienen un contrato con Pigland, Inc., que especifica la entrega de 800 libras de productos de puerco cada lunes. Cada hot dog requie-

re $\frac{1}{4}$ de libra de producto de puerco. Se cuenta con suficiente cantidad del resto de los ingredientes de ambos productos. Por último, la mano de obra consiste en 5 empleados de tiempo completo (40 horas por semana). Cada hot dog requiere 3 minutos de trabajo y cada pan 2 minutos de este insumo. Cada hot dog proporciona una ganancia de \$0.80 y cada pan \$0.30.

Weenies and Buns desea saber cuántos hot dogs y cuántos panes debe producir cada semana para lograr la ganancia más alta posible.

a) Formule un modelo de programación lineal para este problema.

D,I ***b)*** Use el método gráfico para resolver el modelo.

3.1-11.* La compañía manufacturera Omega discontinuó la producción de cierta línea de productos no redituable. Esta medida creó un exceso considerable de capacidad de producción. La administración quiere dedicar esta capacidad a uno o más de tres productos, llamados 1, 2 y 3. En la siguiente tabla se resume la capacidad disponible de cada máquina que puede limitar la producción:

Tipo de máquina	Tiempo disponible (en horas-máquina por semana)
Fresadora	500
Torno	350
Rectificadora	150

El número de horas-máquina que se requieren para elaborar cada unidad de los productos respectivos es

Coeficiente de productividad (en horas-máquina por unidad)

Tipo de máquina	Producto 1	Producto 2	Producto 3
Fresadora	9	3	5
Torno	5	4	0
Rectificadora	3	0	3

El departamento de ventas indica que las ventas potenciales de los productos 1 y 2 exceden la tasa máxima de producción y que las ventas potenciales del producto 3 son de 20 unidades por semana. La ganancia unitaria sería de \$50, \$20 y \$25, para los productos 1, 2 y 3, respectivamente. El objetivo es determinar cuántos productos de cada tipo debe producir la compañía para maximizar la ganancia.

a) Formule un modelo de programación lineal para este problema.

C ***b)*** Utilice una computadora para resolver este modelo mediante el método símplex.

D **3.1-12.** Considere el siguiente problema, donde el valor de c_1 todavía no se ha establecido.

Maximizar $Z = c_1x_1 + x_2$,

sujeta a

$$x_1 + x_2 \leq 6$$
$$x_1 + 2x_2 \leq 10$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

Use el método gráfico para determinar la(s) solución(es) óptima(s) de (x_1, x_2) para los diferentes valores posibles de $c_1 (-\infty < c_1 < \infty)$.

D **3.1-13.** Considere el siguiente problema, donde el valor de k todavía no se ha establecido.

Maximizar $Z = x_1 + 2x_2$,

sujeta a

$$-x_1 + x_2 \leq 2$$
$$x_2 \leq 3$$
$$kx_1 + x_2 \leq 2k + 3, \quad \text{donde } k \geq 0$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

La solución que se usa por ahora es $x_1 = 2, x_2 = 3$. Utilice el análisis gráfico para determinar los valores de k tales que esta solución sea de hecho óptima.

D **3.1-14.** Considere el siguiente problema para el que no se han determinado los valores de c_1 y c_2.

Maximizar $Z = c_1x_1 + c_2x_2$,

sujeta a

$$2x_1 + x_2 \leq 11$$
$$-x_1 + 2x_2 \leq 2$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

Utilice el análisis gráfico para determinar la o las soluciones óptimas de (x_1, x_2) para los diferentes valores posibles de c_1 y c_2. (*Sugerencia:* Separe los casos en los cuales $c_2 = 0$, $c_2 > 0$ y $c_2 < 0$. En los dos últimos casos centre su atención en la razón de c_1 sobre c_2.)

3.2-1. La siguiente tabla resume los hechos importantes sobre dos productos, A y B y los recursos Q, R y S que se requieren para producirlos.

Recurso	Recursos utilizados por unidad de producto: Producto A	Producto B	Cantidad de recursos disponibles
Q	2	1	2
R	1	2	2
S	3	3	4
Ganancia por unidad	3	2	

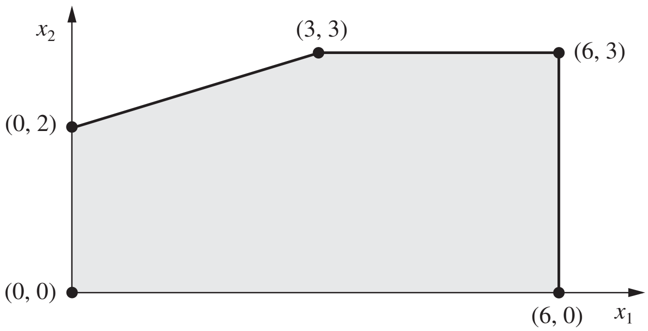
Todos los supuestos de programación lineal se cumplen.

a) Formule un modelo de programación lineal para este problema.

D,I ***b)*** Resuelva este modelo en forma gráfica.

c) Verifique el valor exacto de la solución óptima en *b*) mediante la solución algebraica simultánea de las dos ecuaciones relevantes.

3.2-2. El área sombreada de la siguiente gráfica representa la región factible de un problema de programación lineal cuya función objetivo debe maximizarse.



Diga si cada una de las siguientes afirmaciones es falsa o verdadera y después justifique su respuesta con base en el método gráfico. En cada caso dé un ejemplo de una función objetivo que ilustre su respuesta.

a) Si (3, 3) produce un valor más grande de la función objetivo que (0, 2) y (6, 3), entonces (3, 3) debe ser una solución óptima.
b) Si (3, 3) es una solución óptima y existen soluciones óptimas múltiples, entonces uno de los dos, (0, 2) o (6, 3), también debe ser una solución óptima.
c) El punto (0, 0) no puede ser una solución óptima.

3.2-3.* Hoy es su día de suerte. Acaba de ganar un premio de \$10 000. Dedicará \$4 000 a impuestos y diversiones, pero ha decidido invertir los otros \$6 000. Al oír esta noticia, dos amigos le han ofrecido una oportunidad de convertirse en socio en dos empresas distintas, cada una planeada por uno de ellos. En ambos casos, la inversión incluye dedicar parte de su tiempo el siguiente verano y dinero en efectivo. Para ser un socio *pleno* en el caso del primer amigo debe invertir \$5 000 y 400 horas, y su ganancia estimada (sin tomar en cuenta el valor de su tiempo) sería de \$4 500. Las cifras correspondientes para el segundo caso son \$4 000 y 500 horas, con una ganancia estimada igual a la anterior. Sin embargo, ambos amigos son flexibles y le permitirían asociarse con cualquier *fracción* de participación que quiera. Si elige una participación parcial, todas las cifras dadas para la sociedad plena (inversión de dinero y tiempo, y la ganancia) se pueden multiplicar por esta fracción.

Como de todas formas usted busca un trabajo de verano interesante (máximo 600 horas), ha decidido participar en una o ambas empresas en alguna combinación que maximice su ganancia total estimada. Usted debe resolver el problema de encontrar la mejor combinación.

a) Describa la analogía entre este problema y el de la Wyndor Glass Co. que se presentó en la sección 3.1. Después construya y llene una tabla como la 3.1 para manejar este problema, e identifique las actividades y los recursos.
b) Formule un modelo de programación lineal para este problema.
D,I ***c***) Use el método gráfico para resolver el modelo. ¿Cuál es su ganancia total estimada?

D,I **3.2-4.** Use el método gráfico para encontrar todas las soluciones óptimas del siguiente modelo:

Maximizar $Z = 500x_1 + 300x_2$,

sujeta a

$$\begin{aligned} 15x_1 + 5x_2 &\le 300 \\ 10x_1 + 6x_2 &\le 240 \\ 8x_1 + 12x_2 &\le 450 \end{aligned}$$

y

$$x_1 \ge 0, \quad x_2 \ge 0.$$

D **3.2-5.** Utilice el método gráfico para demostrar que el siguiente modelo no tiene soluciones factibles.

Maximizar $Z = 5x_1 + 7x_2$,

sujeta a

$$\begin{aligned} 2x_1 - x_2 &\le -1 \\ -x_1 + 2x_2 &\le -1 \end{aligned}$$

y

$$x_1 \ge 0, \quad x_2 \ge 0.$$

D **3.2-6.** Suponga que se proporcionaron las siguientes restricciones de un modelo de programación lineal.

$$\begin{aligned} -x_1 + 2x_2 &\le 50 \\ -2x_1 + x_2 &\le 50 \end{aligned}$$

y

$$x_1 \ge 0, \quad x_2 \ge 0.$$

a) Demuestre que la región factible no está acotada.
b) Si el objetivo es maximizar $Z = -x_1 + x_2$, ¿tiene el modelo una solución óptima? Si es así, encuéntrela. Si no, explique las razones de ello.
c) Repita el inciso *b*) cuando el objetivo es maximizar $Z = x_1 - x_2$.
d) En las funciones objetivo con las cuales el modelo no tiene solución óptima, ¿significa que no existen buenas soluciones según el modelo? Explique. ¿Qué es probable que esté mal en la formulación del modelo?

3.3-1. Reconsidere el problema 3.2-3. Indique por qué cada uno de los cuatro supuestos de programación lineal (sección 3.3) parece que se satisface de modo razonable. ¿Es algún supuesto más dudoso que los otros? Si es así, ¿qué debe hacerse para tomar esto en cuenta?

3.3-2. Considere un problema con dos variables de decisión, x_1 y x_2, que representan los niveles de las actividades 1 y 2, respectivamente. Los valores permitidos para cada variable son 0, 1 y 2, en donde las combinaciones factibles de estos valores de las dos variables están determinadas por una serie de restricciones. El objetivo es maximizar cierta medida de desempeño denotada por *Z*. Se estima que los valores de *Z* para los valores posiblemente factibles de (x_1, x_2) son los que se dan en la siguiente tabla:

	x_2		
x_1	0	1	2
0	0	4	8
1	3	8	13
2	6	12	18

Con base en esta información indique si este problema satisface por completo cada uno de los supuestos de programación lineal. Justifique sus respuestas.

3.4-1. Lea el artículo mencionado que describe todo el estudio de IO que se resume en el primer recuadro de aplicación de la sección 3.4. Describa de manera breve cómo se aplicó la programación lineal en este estudio. Después enumere los diferentes beneficios financieros y de otro tipo que resultaron de este estudio.

3.4-2. Lea el artículo referenciado que describe todo el estudio de IO que se resume en el segundo recuadro de aplicación de la sección 3.4. Describa de manera breve cómo se aplicó la programación lineal en este estudio. Después enumere los diferentes beneficios financieros y no financieros que resultaron de este estudio.

3.4-3.* Para cada uno de los cuatro supuestos de programación lineal de la sección 3.3 escriba un análisis de un párrafo sobre la medida en que se aplican correctamente a los siguientes ejemplos que se presentan en la sección 3.4:

a) Diseño de terapia de radiación (Mary).
b) Planeación regional (Southern Confederation of Kibbutzim).
c) Control de contaminación del aire (Nori & Leets Co.).

3.4-4. Para cada uno de los cuatro supuestos de programación lineal de la sección 3.3 escriba un análisis de un párrafo sobre la medida en que se aplican correctamente a los siguientes ejemplos que se presentan en la sección 3.4.

a) Reciclado de desechos sólidos (Save-It Co.).
b) Programación de personal (Union Airways).
c) Distribución de bienes a través de una red (Distribution Unlimited Co.).

D,I **3.4-5.** Utilice el método gráfico para resolver el problema:

Minimizar $Z = 15x_1 + 20x_2,$

sujeta a

$$x_1 + 2x_2 \geq 10$$
$$2x_1 - 3x_2 \leq 6$$
$$x_1 + x_2 \geq 6$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

D,I **3.4-6.** Utilice el método gráfico para resolver este problema:

Minimizar $Z = 3x_1 + 2x_2,$

sujeta a

$$x_1 + 2x_2 \leq 12$$
$$2x_1 + 3x_2 = 12$$
$$2x_1 + x_2 \geq 8$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

D **3.4-7.** Considere el siguiente problema, donde el valor de c_1 no se ha determinado.

Maximizar $Z = c_1x_1 + 2x_2,$

sujeta a

$$4x_1 + x_2 \leq 12$$
$$x_1 - x_2 \geq 2$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

Use el método gráfico para determinar la(s) solución(es) de (x_1, x_2) para los valores posibles de c_1.

D,I **3.4-8.** Considere el siguiente modelo:

Minimizar $Z = 40x_1 + 50x_2,$

sujeta a

$$2x_1 + 3x_2 \geq 30$$
$$x_1 + x_2 \geq 12$$
$$2x_1 + x_2 \geq 20$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

a) Use el método gráfico para resolver este modelo.
b) ¿Cómo varía la solución óptima si la función objetivo cambia a $Z = 40x_1 + 70x_2$? (Puede resultar útil emplear el procedimiento de análisis gráfico y análisis de sensibilidad mediante el IOR Tutorial.)
c) ¿Cómo varía la solución óptima si la tercera restricción funcional cambia a $2x_1 + x_2 \geq 15$? (Puede resultar útil emplear el procedimiento de análisis gráfico y análisis de sensibilidad mediante el IOR Tutorial.)

3.4-9. La carne con papas es el plato favorito de Ralph Edmund. Por eso decidió hacer una dieta continua de sólo estos dos alimentos (más algunos líquidos y suplementos de vitaminas) en todas sus comidas. Ralph sabe que ésa no es la dieta más sana y quiere asegurarse de que toma las cantidades adecuadas de los dos alimentos para satisfacer los requerimientos nutricionales. Él ha obtenido la información nutricional y de costo que se muestra en el siguiente cuadro.

Ralph quiere determinar el número de porciones diarias (pueden ser fraccionales) de res y papas que cumplirían con estos requerimientos a un costo mínimo.

a) Formule un modelo de programación lineal.
D,I ***b)*** Use el método gráfico para resolver el modelo.
C ***c)*** Utilice una computadora para resolver este modelo por el método símplex.

	Gramos de ingrediente por porción		
Ingrediente	**Res**	**Papas**	**Requerimiento diario (gramos)**
Carbohidratos	5	15	≥ 50
Proteínas	20	5	≥ 40
Grasa	15	2	≤ 60
Costo por porción	\$4	\$2	

3.4-10. Web Mercantile vende muchos productos para el hogar mediante un catálogo en línea. La compañía necesita un gran espacio para almacenar los productos. En la actualidad planea rentar espacio para los siguientes 5 meses. Se sabe cuánto espacio necesitará cada mes, pero como dicha superficie es muy variable, puede ser más económico rentar sólo la cantidad necesaria cada mes con contratos

mensuales. Por otro lado, el costo adicional de rentar espacio para meses adicionales es menor que para el primero, y puede ser menos costoso rentar el espacio máximo los 5 meses. Otra opción es el enfoque intermedio de cambiar la cantidad total de espacio rentado (con un nuevo contrato y/o la terminación del anterior) al menos una vez pero no cada mes.

El espacio que se requiere y los costos de los periodos de arrendamiento son los siguientes:

Mes	Espacio requerido (ft^2)
1	30 000
2	20 000
3	40 000
4	10 000
5	50 000

Periodo de arrendamiento (meses)	Costo por ft^2 arrendado
1	$ 65
2	$100
3	$135
4	$160
5	$190

El objetivo es minimizar el costo total de arrendamiento para cumplir con los requerimientos.

a) Formule un modelo de programación lineal para este problema.

C ***b***) Resuelva este modelo por el método símplex.

3.4-11. Larry Edison es el director del centro de cómputo de Buckly College, en donde debe programar las horas de trabajo del personal del centro. Abre desde las 8 a.m. hasta la medianoche. Larry estudió el uso del centro en las diferentes horas del día y determinó los siguientes números de asesores en computación necesarios:

Horarios	Número mínimo de asesores requeridos
8 a.m.-12 p.m.	4
12 p.m.-4 p.m.	8
4 p.m.-8 p.m.	10
8 p.m.-12 a.m.	6

Puede contratar dos tipos de asesores: de tiempo completo y de tiempo parcial. Los primeros trabajan 8 horas consecutivas en cualquiera de los siguientes turnos: matutino (8 a.m.-4 p.m.), vespertino (12 p.m.-8 p.m.) y nocturno (4 p.m.-12 a.m.). Estos asesores ganan $40 por hora.

Los asesores de tiempo parcial pueden trabajar cualquiera de los cuatro turnos enumerados en la tabla anterior y ganan $30 por hora.

Un requisito adicional es que durante todos los periodos debe haber al menos dos asesores de tiempo completo por cada uno de tiempo parcial.

Larry desea determinar cuántos asesores de tiempo completo y cuántos de tiempo parcial debe haber en cada turno para cumplir con los requisitos a un costo mínimo.

a) Formule un modelo de programación lineal para este problema.

C ***b***) Resuelva este modelo por el método símplex.

3.4-12.* La Medequip Company produce equipos de precisión de diagnóstico médico en dos fábricas. Se han recibido pedidos de tres centros médicos para la producción de este mes. La tabla presenta el costo unitario de envío desde cada fábrica a cada centro. Además, muestra el número de unidades que se producirán en cada fábrica y el número de unidades ordenadas por cada cliente.

De \ A	Costo unitario de envío: Cliente 1	Cliente 2	Cliente 3	Producción
Fábrica 1	$600	$800	$700	400 unidades
Fábrica 2	$400	$900	$600	500 unidades
Tamaño de unidades	300 unidades	200 unidades	400 unidades	

Ahora debe tomar la decisión sobre el plan de cuántas unidades enviar de cada fábrica a cada cliente.

a) Formule un modelo de programación lineal.

C ***b***) Resuelva el modelo por el método símplex.

3.4-13.* Al Ferris tiene $60 000 que desea invertir ahora para usar lo que se acumule en la compra de un fondo de retiro en 5 años. Después de consultar a su asesor financiero, le ofrecieron cuatro tipos de inversiones de ingreso fijo, las inversiones A, B, C y D.

Las inversiones A y B están disponibles al principio cada uno de los siguientes 5 años (años 1 a 5). Cada dólar invertido en A al iniciar el año reditúa $1.40 (ganancia de $0.40) 2 años después (a tiempo para invertir de inmediato). Cada dólar invertido en B al principio de un año ofrece $1.70 tres años después.

Las inversiones C y D estarán disponibles una sola vez en el futuro. Cada dólar invertido en C al principio del año 2 genera $1.90 al final del 5. Cada dólar invertido en D al principio del año 5 produce $1.30 al final de ese año.

Al desea saber cuál plan de inversión maximiza la cantidad de dinero acumulada al principio del año 6.

a) Todas las restricciones funcionales de este problema se pueden expresar como igualdades. Para hacerlo, sean A_t, B_t, C_t y D_t las cantidades respectivas invertidas en A, B, C y D al principio del año *t* si la inversión está disponible y llegará a su madurez al final del año 5. También sea R_t la cantidad de dinero disponible *no* invertida al principio del año *t* (también disponible para invertir en un año posterior). Así, la cantidad invertida al principio del año *t más* R_t debe ser igual a la cantidad disponible para inversión en ese tiempo. Escriba una ecuación en términos de las variables relevantes para el inicio de cada uno de los 5 años para obtener las cinco restricciones funcionales de este problema.

b) Formule un modelo de programación lineal completo.

C ***c***) Resuelva este modelo por el método símplex.

3.4-14. Metalco Company desea hacer una nueva aleación con 40% de aluminio, 35% de zinc y 25% de plomo a partir de varias aleaciones disponibles que tienen las siguientes propiedades:

Propiedad	Aleación 1	2	3	4	5
Porcentaje de aluminio	60	25	45	20	50
Porcentaje de zinc	10	15	45	50	40
Porcentaje de plomo	30	60	10	30	10
Costo ($/libra)	77	70	88	84	94

El objetivo es determinar las proporciones de estas aleaciones que deben mezclarse para producir la nueva aleación a un costo mínimo.

a) Formule un modelo de programación lineal.
C *b*) Resuelva este modelo por el método símplex.

3.4-15.* Un avión de carga tiene tres compartimientos para almacenar: delantero, central y trasero. Estos compartimientos tienen un límite de capacidad tanto de *peso* como de *espacio*. Los datos se resumen a continuación:

Compartimiento	Capacidad de peso (ton)	Capacidad de espacio (ft^3)
Delantero	12	7 000
Central	18	9 000
Trasero	10	5 000

Más aún, para mantener el avión balanceado, el peso de la carga en los respectivos compartimientos debe ser proporcional a su capacidad.

Se tienen ofertas para transportar cuatro cargamentos en un vuelo próximo ya que se cuenta con espacio:

Carga	Peso (ton)	Volumen (ft^3/ton)	Ganancia ($/ton)
1	20	500	320
2	16	700	400
3	25	600	360
4	13	400	290

Se puede aceptar cualquier fracción de estas cargas. El objetivo es determinar cuál cantidad de cada carga debe aceptarse (si se acepta) y cómo distribuirla en los compartimientos para maximizar la ganancia del vuelo.
a) Formule un modelo de programación lineal.
C *b*) Resuelva el modelo por el método símplex para encontrar una de sus soluciones óptimas múltiples.

3.4-16. Oxbridge University tiene una computadora grande para uso de académicos, estudiantes de doctorado y ayudantes de investigación. Durante las horas hábiles debe haber un trabajador para operar y dar mantenimiento a la computadora y realizar algunos servicios de programación. Beryl Ingram, director del centro de cómputo, coordina la operación.

Al principio del semestre de otoño, Beryl se enfrenta al problema de asignar horas de trabajo distintas a sus operadores. Debido a que éstos son estudiantes de la universidad, están disponibles para el trabajo sólo un número limitado de horas al día, como se muestra en la tabla.

Operadores	Tasa salarial	Máximo de horas disponibles				
		Lun.	Mar.	Mier.	Jue.	Vie.
K. C.	$10.00/hora	6	0	6	0	6
D. H.	$10.10/hora	0	6	0	6	0
H. B.	$ 9.90/hora	4	8	4	0	4
S. C.	$ 9.80/hora	5	5	5	0	5
K. S.	$10.80/hora	3	0	3	8	0
N. K.	$11.30/hora	0	0	0	6	2

Hay seis operadores (cuatro de licenciatura y dos de posgrado). Todos tienen salarios diferentes según su experiencia con computadoras y su aptitud para programar. La tabla muestra estos salarios junto con el número máximo de horas al día que cada uno puede trabajar.

Se garantiza a cada operador un número mínimo de horas de trabajo a la semana que lo mantendrán con un conocimiento adecuado de la operación. Este nivel se estableció de modo arbitrario en 8 horas por semana para licenciatura (K. C., D. H., H. B. y S. C.) y 7 horas por semana para posgrado (K. S. y N. K).

El centro de cómputo debe abrir de 8 a.m. a 10 p.m. de lunes a viernes con un operador de guardia en este horario. Sábados y domingos, lo operan otras personas.

Debido al presupuesto reducido, Beryl tiene que minimizar el costo. Por lo tanto, quiere determinar el número de horas que debe asignar a cada operador cada día.
a) Formule un modelo de programación lineal para este problema.
C *b*) Resuelva este modelo por el método símplex.

3.4-17. Joyce y Marvin tienen una guardería. Intentan decidir qué dar a los niños de almuerzo. Desean mantener sus costos bajos, pero también deben cumplir con los requerimientos nutritivos de los niños. Ya decidieron darles sándwiches de mantequilla de maní y mermelada y alguna combinación de galletas, leche y jugo de naranja. El contenido nutritivo de cada alimento y su costo se presenta en la siguiente tabla.

Ingredientes	Calorías de grasa	Calorías totales	Vitamina C (mg)	Proteína (g)	Costo (¢)
Pan (l rebanada)	10	70	0	3	5
Mantequilla de maní (1 cuch.)	75	100	0	4	4
Mermelada de fresa (1 cuch.)	0	50	3	0	7
Galleta (1 pieza)	20	60	0	1	8
Leche (1 taza)	70	150	2	8	15
Jugo (1 taza)	0	100	120	1	35

Los requerimientos nutritivos son los siguientes. Cada niño debe recibir de 400 a 600 calorías. No más de 30% de las calorías totales deben provenir de grasas. Cada niño debe consumir al menos 60 mg de vitamina C y 12 g de proteína. Todavía más, por razones prácticas, cada niño necesita 2 rebanadas de pan (para un sándwich), al menos el doble de mantequilla de maní que de mermelada y al menos una tasa de líquido (leche y/o jugo de naranja).

Joyce y Marvin desean seleccionar las opciones de alimento para cada niño que minimice el costo mientras cumple con los requerimientos establecidos.
a) Formule un modelo de programación lineal para este problema.
C *b*) Resuelva el modelo por el método símplex.

3.5-1. Lea el artículo referenciado que describe el estudio de IO que se resume en el Recuadro de aplicación de la sección 3.5. Describa brevemente cómo se aplicó la programación lineal en este estudio. Después enumere los diferentes beneficios financieros y de otro tipo que resultaron de este estudio.

3.5-2.* Se cuenta con los siguientes datos de un problema de programación lineal cuyo objetivo es maximizar la ganancia de asignar tres recursos a dos actividades no negativas.

Recurso	Uso de recursos por unidad de cada actividad		Cantidad de recursos disponible
	Actividad 1	Actividad 2	
1	2	1	10
2	3	3	20
3	2	4	20
Contribución por unidad	$20	$30	

Contribución por unidad = ganancia por unidad de la actividad.

a) Formule un modelo de programación lineal para este problema.
D,I ***b***) Use el método gráfico para resolver este modelo.
c) Despliegue el modelo en una hoja de cálculo.
d) Use la hoja de cálculo para verificar las siguientes soluciones: $(x_1, x_2) = (2, 2), (3, 3), (2, 4), (4, 2), (3, 4), (4, 3)$. ¿Cuáles son factibles? ¿Cuál de las soluciones factibles tiene el mejor valor de la función objetivo?
C ***e***) Utilice el Excel Solver para resolver el modelo por el método símplex.

3.5-3. Ed Butler es gerente de producción de Bilco Corporation, que produce tres tipos de refacciones para automóviles. La manufactura de cada parte requiere procesamiento en dos máquinas, con los siguientes tiempos de procesado (en horas):

Máquina	Parte		
	A	B	C
1	0.02	0.03	0.05
2	0.05	0.02	0.04

Cada máquina está disponible 40 horas al mes. La ganancia unitaria de cada parte fabricada está dada por:

	Parte		
	A	B	C
Utilidad	$300	$250	$200

Ed quiere determinar la mezcla de refacciones que debe producir para maximizar la ganancia total.
a) Formule un modelo de programación lineal para este problema.
b) Despliegue el modelo en una hoja de Excel.
c) Realice tres estimaciones de la solución óptima. Use la hoja de cálculo para verificar la factibilidad de cada una y, si es posible, encuentre el valor de la función objetivo. ¿Qué estimación tiene el mejor valor de la función objetivo?
d) Utilice Excel Solver para resolver este modelo por el método símplex.

3.5-4. Usted cuenta con los siguientes datos de un problema de programación lineal cuyo objetivo es minimizar el costo de realizar dos actividades no negativas para lograr tres beneficios que nunca estén por debajo de ciertos niveles mínimos.

Beneficio	Contribución al beneficio por unidad de actividad		Nivel mínimo aceptable
	Actividad 1	Actividad 2	
1	5	3	60
2	2	2	30
3	7	9	126
Costo unitario	$60	$50	

a) Formule un modelo de programación lineal para este problema.
D,I ***b***) Utilice el método gráfico para resolver este modelo.
c) Despliegue el modelo en una hoja de Excel.
d) Use la hoja de cálculo para verificar las siguientes soluciones: $(x_1, x_2) = (7, 7), (7, 8), (8, 7), (8, 8), (8, 9), (9, 8)$. ¿Cuáles soluciones son factibles? ¿Cuál de ellas tiene el mejor valor objetivo?
C ***e***) Use el Excel Solver para resolver el modelo por el método símplex.

3.5-5.* Fred Jonasson administra la granja de su familia. Para complementar varios alimentos que se cultivan en la granja, Fred también cría cerdos para venta y desea determinar las cantidades de los distintos tipos de alimento disponibles (maíz, grasas y alfalfa) que debe dar a cada cerdo. Como éstos se comerán cualquier mezcla de estos tipos de alimento, el objetivo es determinar cuál de ellas cumple ciertos requisitos nutritivos a un *costo mínimo*. En la siguiente tabla se presentan las unidades de cada tipo de ingrediente nutritivo básico que contiene 1 kilogramo de cada tipo de alimento, junto con los requisitos de nutrición diarios y los costos de los alimentos:

Ingrediente nutritivo	Kilogramo de maíz	Kilogramo de nutrimento	Kilogramo de alfalfa	Requerimiento mínimo diario
Carbohidratos	90	20	40	200
Proteína	30	80	60	180
Vitaminas	10	20	60	150
Costo (¢)	84	72	60	

a) Formule el modelo de programación lineal para este problema.
b) Despliegue el modelo en una hoja de Excel.
c) Utilice la hoja de cálculo para verificar: si $(x_1, x_2, x_3) = (1, 2, 2)$ es factible y, si lo es, cuál sería el costo diario de esta dieta. ¿Cuántas unidades de cada ingrediente nutritivo proporciona al día esta dieta?
d) Tome unos minutos para usar un enfoque de prueba y error con la hoja de cálculo a fin de obtener la mejor estimación de la solución óptima. ¿Cuál es el costo diario de su solución?
C ***e***) Use el Excel Solver para resolver el modelo por el método símplex.

3.5-6. Maureen Laird es directora de inversiones de Alva Electric Co., empresa importante en el medio oeste. La compañía ha programado la construcción de nuevas plantas hidroeléctricas a 5, 10 y 20 años para cumplir con las necesidades de la creciente población en la región que sirve. Maureen debe invertir parte del dinero de la compañía para cubrir sus necesidades de efectivo futuras. Puede comprar sólo tres tipos de activos, cada una de las cuales cuesta 1 millón. Se

pueden comprar unidades fraccionarias. Los activos producen ingresos a 5, 10 y 20 años, y el ingreso se necesita para cubrir necesidades mínimas de flujos de efectivo en esos años. (Cualquier ingreso arriba del mínimo que se requiere para cada periodo se usará para incrementar el pago de dividendos a los accionistas en lugar de ahorrarlo para ayudar a cumplir con los requerimientos mínimos de efectivo del siguiente periodo.) La tabla que se presenta a continuación muestra la cantidad de ingreso generada por cada unidad de acciones y la cantidad mínima de ingreso requerida para cada periodo futuro en que se construirá una nueva planta.

	Ingresos por acción			Flujo de efectivo requerido
Año	**Activo 1**	**Activo 2**	**Activo 3**	
5	$2 millones	$1 millón	$0.5 millones	$400 millones
10	$0.5 millones	$0.5 millones	$1 millón	$100 millones
20	0	$1.5 millones	$2 millones	$300 millones

Maureen desea determinar la mezcla de inversiones en estas acciones que cubrirá los requerimientos de efectivo y que minimizará la cantidad total invertida.

a) Formule un modelo de programación lineal para este problema.
b) Despliegue el modelo en una hoja de cálculo.
c) Utilice la hoja de cálculo para verificar la posibilidad de comprar 100 unidades de la acción 1, 100 de la acción 2 y 200 de la 3. ¿Cuánto efectivo generará esta mezcla de inversiones dentro de 5, 10 y 20 años, respectivamente? ¿Cuál será la cantidad total invertida?
d) Utilice el enfoque de prueba y error con la hoja de cálculo para obtener su mejor solución óptima. ¿Cuál es la inversión total de su solución?
C ***e***) Use Excel Solver para resolver el modelo por el método símplex.

3.6-1. La Philbrick Company tiene dos plantas en lados opuestos de Estados Unidos. Cada una produce los mismos dos productos y los vende a distribuidores en su mitad del país. Ya se recibieron las órdenes de los distribuidores para los próximos 2 meses (febrero y marzo); el número de unidades que se requieren se muestra en la tabla. (La compañía no está obligada a cumplir totalmente estas órdenes, pero lo hará, si puede, sin disminuir sus ganancias.)

	Planta 1		Planta 2	
Producto	**Febrero**	**Marzo**	**Febrero**	**Marzo**
1	3 600	6 300	4 900	4 200
2	4 500	5 400	5 100	6 000

Cada planta tiene 20 días de producción disponibles en febrero y 23 en marzo para producir y enviar los productos. Los inventarios se agotan al final de enero, pero cada planta tiene suficiente capacidad de inventario para 1 000 unidades en total de los dos productos, si se produce un exceso en febrero para venta en marzo. En cualquier planta, el costo de mantener inventario de esta manera es de $3 por unidad del producto 1 y $4 por unidad del producto 2.

Cada planta tiene los mismos dos procesos de producción que se pueden usar para producir cualquiera de estos productos. El costo de producción por unidad producida se muestra en la tabla para cada proceso en cada planta.

	Planta 1		Planta 2	
Producto	**Proceso 1**	**Proceso 2**	**Proceso 1**	**Proceso 2**
1	$62	$59	$61	$65
2	$78	$85	$89	$86

A continuación se presenta la tasa de producción de cada producto (número de unidades de ese producto fabricadas por día) mediante cada proceso en cada planta.

	Planta 1		Planta 2	
Producto	**Proceso 1**	**Proceso 2**	**Proceso 1**	**Proceso 2**
1	100	140	130	110
2	120	150	160	130

El ingreso neto por ventas (precio de venta menos costos de envío normal) que recibe la compañía cuando una planta vende los productos a sus propios clientes (distribuidores en su mitad del país) es de $83 por unidad del producto 1 y $112 por unidad del producto 2. Sin embargo, también es posible (y en ocasiones deseable) que una planta haga un envío a la otra mitad del país para ayudar a satisfacer la venta de la otra. Cuando esto ocurre se incurre en un costo adicional de $9 en el caso del producto 1 y $7 en el del producto 2.

La administración debe determinar cuánto fabricar de cada producto mediante cada proceso en cada planta cada mes, al igual que cuánto debe vender cada planta de cada producto cada mes y cuánto debe enviar cada planta de cada producto cada mes a los clientes de la otra planta. El objetivo es determinar el plan factible que maximice la ganancia total (ingresos netos por venta menos la suma de los costos de producción, de inventario y los costos adicionales de envío).

a) Formule un modelo completo de programación lineal en forma algebraica para mostrar las restricciones individuales y las variables de decisión.
C ***b***) Formule el mismo modelo en una hoja de Excel. Después use Excel Solver para resolverlo.
C ***c***) Use MPL para formular el modelo en forma compacta. Después use CPLEX de MPL para resolverlo.
C ***d***) Use LINGO para formular el modelo en forma compacta. Después use LINGO para resolverlo.

C **3.6-2.** Reconsidere el problema 3.1-11.
a) Use MPL/CPLEX para formular y resolver el modelo para este problema.
b) Use LINGO para formular y resolver este modelo.

C **3.6-3.** Reconsidere el problema 3.4-12.
a) Use MPL/CPLEX para formular y resolver el modelo para este problema.
b) Use LINGO para formular y resolver este modelo.

C **3.6-4.** Reconsidere el problema 3.4-16.
a) Use MPL/CPLEX para formular y resolver el modelo para este problema.
b) Use LINGO para formular y resolver este modelo.

C **3.6-5.** Reconsidere el problema 3.5-5.

a) Use MPL/CPLEX para formular y resolver el modelo para este problema.

b) Use LINGO para formular y resolver este modelo.

C **3.6-6.** Reconsidere el problema 3.5-6.

a) Use MPL/CPLEX para formular y resolver el modelo para este problema.

b) Use LINGO para formular y resolver este modelo.

3.6-7. Una fábrica grande de papel, la Quality Paper Corporation, tiene 10 molinos de papel para surtir a 1 000 clientes. Usa tres tipos alternativos de máquinas y cuatro tipos de materia prima para hacer cinco tipos diferentes de papel. Por lo tanto, la compañía debe desarrollar un plan detallado para distribuir mensualmente la producción, con el objeto de minimizar el costo total de producir y distribuir el papel durante el mes. En particular, es necesario determinar conjuntamente la cantidad de cada tipo de papel que debe producir en cada planta, en cada tipo de máquina *y* la cantidad de cada tipo de papel que debe enviar de cada planta a cada cliente.

Los datos relevantes se pueden expresar de manera simbólica como sigue:

D_{jk} = número de unidades del tipo de papel k demandadas por el cliente j,

r_{klm} = número de unidades de materia prima m necesarias para producir 1 unidad del tipo de papel k en la máquina tipo l,

R_{im} = número de unidades de materia prima m disponibles en la planta i,

c_{kl} = número de unidades de capacidad de la máquina tipo l que producirán una unidad de papel tipo k,

C_{il} = número de unidades de capacidad de la máquina tipo l disponibles en la planta i,

P_{ikl} = costo de producción de cada unidad de papel tipo k producida en la máquina tipo l en la planta i,

T_{ijk} = costo de transporte de cada unidad de papel tipo k enviada de la planta i al cliente j.

a) Utilice estos símbolos para formular a mano un modelo de programación lineal para este problema.

b) ¿Cuántas restricciones funcionales y variables de decisión tiene este modelo?

C ***c***) Use MPL para formular este problema.

C ***d***) Use LINGO para formular este problema.

3.7-1. De la última parte de las referencias seleccionadas que se presentan al final del capítulo, seleccione una de las aplicaciones de programación lineal ganadoras de premios. Lea el artículo y después escriba un resumen de dos páginas de la aplicación y los beneficios que proporcionó (incluya los beneficios no financieros).

3.7-2. De la última parte de las referencias seleccionadas que se presentan al final del capítulo, seleccione tres de las aplicaciones de programación lineal ganadoras de premios. Para cada una, lea el artículo y después escriba un resumen de una página de la aplicación y los beneficios que proporcionó (incluya los beneficios no financieros).

CASOS

CASO 3.1 Ensamble de automóviles

Automobile Alliance, una gran compañía manufacturera de automóviles, organiza los vehículos que fabrica en tres familias: camiones, automóviles pequeños y una familia de autos medianos y de lujo. Una planta fuera de Detroit, MI, ensambla dos modelos de la familia de autos medianos y de lujo. El primer modelo, el Thrillseeker, es un sedán cuatro puertas con asientos de vinil, interiores de plástico, características estándar y un excelente rendimiento. Se promociona como una buena compra para familias de clase media con presupuestos reducidos. Cada Thrillseeker que se vende genera una ganancia modesta de $3 600 para la compañía. El segundo modelo, el Classy Cruiser, es un sedán de lujo de dos puertas con asientos de piel, interiores de madera, características personalizadas y gran capacidad de navegación. Se vende como un símbolo de opulencia a familias de clase media-alta y cada uno genera una buena ganancia de $5 400.

Rachel Rosencrantz, gerente de la planta de ensamblado, debe decidir el programa de producción del próximo mes. En especial, debe determinar cuántos Thrillseekers y cuántos Classy Cruisers se tienen que ensamblar en la planta para maximizar la ganancia de la compañía. Sabe que la planta tiene una capacidad de 48 000 horas de mano de obra al mes. También, que para ensamblar un Thrillseeker se emplean 6 horas-hombre y un Cruise Classy 10.5 horas-hombre.

Debido a que en la planta sólo se ensambla, las partes que se requieren para los dos modelos no se producen en ella. En su lugar, se envían de otras plantas ubicadas en el área de Michigan. Por ejemplo, llantas, volantes, ventanas, asientos y puertas llegan de varias plantas proveedoras. Para el próximo mes, Rachel sabe que podrá obtener sólo 20 000 puertas (10 000 izquierdas y 10 000 derechas) del proveedor de ellas. Una huelga de trabajadores forzó el cierre de esa fábrica durante varios días, y no podrá cumplir con su programa de producción para el siguiente mes. Tanto el Thrillseeker como el Classy Cruiser usan la misma puerta.

Además, un pronóstico reciente de la compañía sobre la demanda del mes de los diferentes modelos sugiere que la venta del Classy Cruiser se limitaría a 3 500 autos. No existe un tope a la demanda del Thrillseeker dentro de los límites de capacidad de la planta de ensamblado.

a) Formule y resuelva un problema de programación lineal para determinar el número de autos Thrillseeker y Classy Cruiser que deben ensamblarse.

Antes de tomar las decisiones de producción finales, Rachel planea explorar los siguientes aspectos por separado, excepto donde se indique otra cosa.

b) El departamento de marketing sabe que puede intentar una campaña de publicidad de $500 000 que elevará la demanda del Classy Cruiser 20% el próximo mes. ¿Debe realizarse la campaña?

c) Rachel sabe que puede aumentar la capacidad de producción de la planta el próximo mes si usa tiempo extra. El incremento de horas-hombre puede ser de 25%. Con la nueva capacidad, ¿cuántos modelos Thrillseeker y cuántos Classy Cruiser deben ensamblarse?

d) Rachel sabe que el tiempo extra genera un costo adicional. ¿Cuál es la máxima cantidad que debe estar dispuesta a pagar por todo el tiempo extra adicional al costo del tiempo normal? Exprese su respuesta como una sola suma.

e) Rachel estudia la opción de usar tanto la campaña de publicidad como las horas de tiempo extra. La campaña eleva 20% la demanda del Classy Cruiser y el tiempo extra aumenta 25% la capacidad de la planta. ¿Cuántos modelos Thrillseeker y cuántos Classy Cruiser deben ensamblarse con la campaña publicitaria y las horas extra si cada Classy Cruiser que se venda mantendrá su contribución de 50% más que la venta de un Thrillseeker?

f) Si se sabe que la campaña de publicidad cuesta $500 000 y el uso máximo de horas-trabajo de tiempo extra cuesta $ 1 600 000 más que el tiempo normal, la solución que se encontró en el inciso *e*) ¿es adecuada comparada con la solución del inciso *a*)?

g) Automobile Alliance ha determinado que, en realidad, los distribuidores hacen grandes descuentos al precio del Thrillseeker para sacarlo del lote. Por un acuerdo de ganancias compartidas con ellos, la compañía no obtendrá la ganancia de $3 600 en el Thrillseeker sino que ganará sólo $2 800. Determine el número de autos Thrillseeker y de Classy Cruiser que deben ensamblarse dado este nuevo precio con descuento.

h) La compañía descubrió problemas de calidad en el Thrillseeker mediante pruebas aplicadas aleatoriamente a unidades del Thrillseeker al final de la línea de ensamblado. Los inspectores detectaron que en más de 60% de los casos, dos de las cuatro puertas del automóvil no sellaban bien. Como el porcentaje de autos Thrillseeker defectuosos determinado por el muestreo aleatorio es tan alto, el supervisor de planta decidió realizar pruebas de control de calidad a todos los vehículos al final de la línea. Debido a las pruebas adicionales, el tiempo para ensamblar un auto aumentó de 6 a 7.5 horas. Determine el número de unidades de cada modelo que deben ensamblarse dado este nuevo tiempo de ensamblado.

i) El consejo directivo de Alliance desea captar un mayor porcentaje de mercado para el sedán de lujo y quisiera cumplir con toda la demanda del Classy Cruiser. Por ello, pidieron a Rachel que determine cuánto disminuiría la ganancia de su planta de ensamblado comparada con la ganancia del inciso *a*). Pueden pedirle que cumpla con toda la demanda de este modelo sólo si la disminución de la ganancia no supera los $2 000 000.

j) Rachel quiere tomar la decisión final combinando todas las consideraciones de los incisos *f*), *g*) y *h*). ¿Cuáles son sus decisiones finales respecto de la campaña publicitaria, las horas extra y el número de autos Thrillseeker y Classy Cruiser que se deben ensamblar?

RESUMEN DE LOS CASOS ADICIONALES EN NUESTRO SITIO WEB (www.mhhe.com/hillier)

Caso 3.2 Disminución de costos en una cafetería

Este caso se enfoca en cierto problema que tiene un interés especial para muchos estudiantes. ¿Cómo puede el administrador de una cafetería universitaria elegir los ingredientes de un platillo para darle un sabor suficientemente bueno para los estudiantes, al mismo tiempo que disminuye sus costos? En este caso se pueden utilizar modelos de programación lineal con sólo dos variables de decisión para abordar siete diferentes aspectos a los que se tiene que enfrentar el administrador.

Caso 3.3 Asignación de personal en un centro de llamadas

El California Children's Hospital ha recibido numerosas quejas de clientes debido a su confuso proceso descentralizado de citas y registro. Por lo tanto, se ha decidido centralizar el proceso mediante un centro de llamadas dedicado sólo a citas y registro. El administrador del hospital debe desarrollar un plan que le ayude a decidir cuántos empleados de cada tipo (de tiempo completo o de medio tiempo, que hablen inglés, español o que sean bilingües) se deben contratar para cada uno de los posibles turnos de trabajo. Se requiere de programación lineal para encontrar un plan que minimice el costo total de proporcionar un nivel de servicio satisfactorio a lo largo de las 14 horas que el centro de llamadas estará abierto todos los días de la semana. El modelo requiere más de dos variables de decisión, por lo cual se necesitará algún paquete de software como los descritos en las secciones 3.5 y 3.6 o en el apéndice 3.1 para resolver las dos versiones del modelo.

Caso 3.4 Promoción de un cereal para el desayuno

El vicepresidente de comercialización de la Super Grain Corporation necesita desarrollar una campaña promocional para el nuevo cereal para el desayuno que ha lanzado su compañía. Se han elegido tres medios de comunicación para la campaña, pero ahora se deben tomar las decisiones acerca de cuánto se utilizará en cada medio. Las restricciones incluyen límites a los presupuestos para publicidad y planeación, el limitado número de anuncios de televisión que se pueden realizar, así como ciertos requerimientos para llegar a dos audiencias objetivo especiales (niños pequeños y sus padres) y para hacer uso completo de un programa de rebajas. El modelo de programación lineal correspondiente requiere más de dos variables de decisión, por lo que se necesitará algún paquete de software como los descritos en las secciones 3.5 y 3.6 o en el apéndice 3.1 para resolver el modelo. En este caso también se pide un análisis para determinar en qué medida satisface el problema los cuatro supuestos de la programación lineal. ¿La programación lineal en realidad proporciona una base racional para tomar la decisión en esta situación? (El caso 12.3 será una continuación de este caso.)

CAPÍTULO 4

Solución de problemas de programación lineal: método símplex

Es el momento de comenzar a estudiar el *método símplex*, un procedimiento general para resolver problemas de programación lineal. Desarrollado por George Dantzig[1] en 1947, se ha comprobado su extraordinaria eficiencia, y se usa en forma rutinaria para resolver problemas grandes en las computadoras de hoy en día. Excepto en el caso de problemas muy pequeños, se ejecuta siempre en una computadora y existe una amplia variedad de paquetes complejos de software para ello. También se usan extensiones y variaciones del método símplex para realizar *análisis posóptimo* (que incluye el análisis de sensibilidad) del modelo.

En este capítulo se describen y ejemplifican las características principales del método símplex. En la primera sección se presenta su naturaleza general junto con su representación geométrica. En las tres secciones subsecuentes se desarrolla el procedimiento para resolver cualquier modelo de programación lineal que se establezca en nuestra forma estándar (maximización, todas las restricciones funcionales de la forma $\leq$ y restricciones de no negatividad sobre todas las variables) y que sólo tenga cantidades *no negativas* en el lado derecho b_i de las restricciones funcionales. En la sección 4.5 se presentan ciertos detalles sobre cómo romper empates, y en la sección 4.6 se describe la adaptación del método símplex a otras formas de modelos. Después se presenta el análisis posóptimo (sección 4.7) y se describe el manejo de este método en computadora (sección 4.8). En la sección 4.9 se introduce una alternativa al método símplex (el enfoque de punto interior) para resolver problemas de programación lineal grandes.

4.1 ESENCIA DEL MÉTODO SÍMPLEX

El método símplex es un procedimiento *algebraico*. Sin embargo, sus conceptos fundamentales son *geométricos*. La comprensión de estos conceptos geométricos proporciona una fuerte intuición sobre la forma en que opera el método símplex y las razones de su elevada eficiencia. Por lo tanto, antes de profundizar en los detalles algebraicos, se dedicará esta sección a enfocar el método desde un punto de vista geométrico.

Para ilustrar los conceptos geométricos generales se usará el ejemplo de la Wyndor Glass Co. de la sección 3.1. (En las secciones 4.2 y 4.3 se usa el *álgebra* del método símplex para resolver este mismo ejemplo.) En la sección 5.1 se profundiza en estos conceptos para resolver problemas grandes.

[1] Ampiamente reconocido como el pionero más importante de la investigación de operaciones, a George Dantzig se le conoce comúnmente como el *padre de la programación lineal* debido al desarrollo del método símplex y a una serie de contribuciones clave subsecuentes. Los autores tuvieron el privilegio de ser sus colegas en el Departamento de Investigación de Operaciones de la Universidad de Stanford por casi 30 años. El doctor Dantzig permaneció profesionalmente activo hasta su fallecimiento en 2005 a la edad de 90 años.

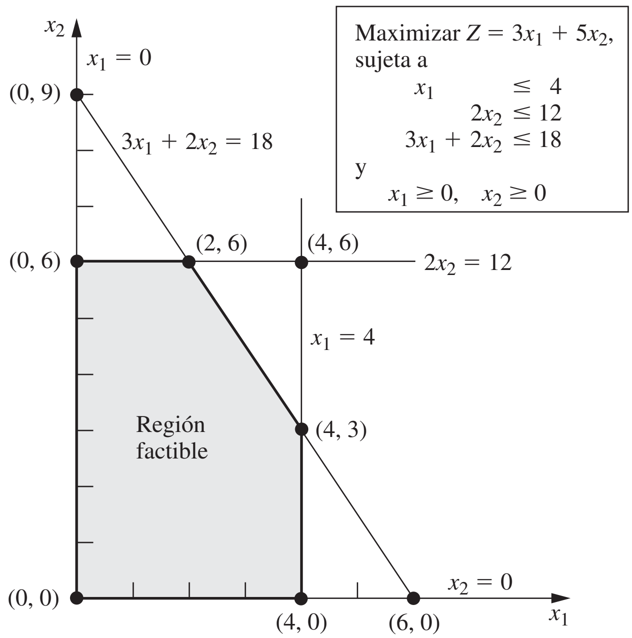


FIGURA 4.1
Restricciones de frontera y soluciones en los vértices del problema de la Wyndor Glass Co.

Para refrescar la memoria, el modelo y la gráfica de este ejemplo se repiten en la figura 4.1. Se marcaron las cinco fronteras de restricción y sus puntos de intersección puesto que son puntos clave para el análisis. Aquí, cada **frontera de restricción** es una recta que marca el límite de lo que permite la restricción correspondiente. Los puntos de intersección son las **soluciones en los vértices** del problema. Los cinco puntos que se encuentran en los vértices de la *región factible* —(0, 0), (0, 6), (2, 6), (4, 3) y (4, 0)— son las *soluciones factibles en los vértices* (**soluciones FEV**). [Los otros tres —(0, 9), (4, 6) y (6, 0)— se llaman *soluciones no factibles en un vértice*.]

En este ejemplo, cada solución en un vértice se encuentra en la intersección de *dos* fronteras de restricción. (En el caso de un problema de programación lineal con n variables de decisión, cada una de sus soluciones en los vértices se encuentra en la intersección de n fronteras de restricciones.)[2] Algunos pares de soluciones FEV de la figura 4.1 comparten una frontera de restricción, y otros no. Será importante distinguir estos casos con la siguiente definición general.

> En cualquier problema de programación lineal con n variables de decisión, dos soluciones FEV son **adyacentes** entre sí cuando comparten $n - 1$ fronteras de restricción. Dos soluciones FEV adyacentes están conectadas por un segmento de recta que se encuentra en estas mismas fronteras de restricción compartidas. Dicho segmento de recta recibe el nombre de **arista** de la región factible.

Como en el ejemplo $n = 2$, dos de sus soluciones FEV son adyacentes si comparten *una* frontera de restricción; por ejemplo, (0, 0) y (0, 6) son adyacentes porque comparten la frontera $x_1 = 0$. La región factible de la figura 4.1 tiene cinco aristas que consisten en los cinco segmentos que forman la frontera de esta región. Observe que de cada solución FEV salen dos aristas. En consecuencia, cada solución FEV tiene dos soluciones FEV adyacentes (cada una se encuentra en el otro punto terminal de una de las dos aristas), como se enumera en la tabla 4.1. (En cada renglón de esta tabla, la solución FEV de la primera columna es adyacente a las dos soluciones FEV de la segunda columna, pero las dos soluciones de esta última *no* son adyacentes entre sí.)

Una razón para analizar las soluciones FEV adyacentes es la siguiente propiedad general de las soluciones, que proporciona una manera muy útil de verificar si una solución FEV es óptima.

[2] Aunque una solución en un vértice está definida en términos de n fronteras de restricciones cuyas intersecciones dan esta solución, también es posible que una o más fronteras adicionales pasen por este mismo punto.

TABLA 4.1 Soluciones FEV adyacentes para cada solución FEV del problema de la Wyndor Glass Co.

Solución FEV	Sus soluciones FEV adyacentes
(0, 0)	(0, 6) y (4, 0)
(0, 6)	(2, 6) y (0, 0)
(2, 6)	(4, 3) y (0, 6)
(4, 3)	(4, 0) y (2, 6)
(4, 0)	(0, 0) y (4, 3)

> **Prueba de optimalidad:** Considere cualquier problema de programación lineal que posea al menos una solución óptima. Si una solución FEV no tiene soluciones FEV *adyacentes* que sean *mejores* (según el valor de Z), entonces ésa *debe* ser una solución *óptima*.

Así, por ejemplo (2, 6) debe ser óptima sólo porque su valor correspondiente de $Z = 36$ es más grande que $Z = 30$ para (0, 6) y $Z = 27$ para (4, 3). (En la sección 5.1 se analizará con mayor profundidad por qué se cumple esta propiedad.) Esta prueba de optimalidad se usa en el método símplex para determinar cuándo se ha llegado a una solución óptima.

En este momento, es posible aplicar el método símplex a un ejemplo.

Solución del ejemplo

Se presenta aquí un bosquejo de cómo utilizar el método símplex (desde el punto de vista geométrico) para resolver el problema de la Wyndor Glass Co. En cada paso, primero se establece la conclusión y después se expone la razón entre paréntesis. (Remítase a la figura 4.1 para tener una imagen gráfica del problema.)

Paso inicial: Elija (0, 0) como la solución FEV *inicial* para examinarla. (Ésta es una elección conveniente porque no se requieren cálculos para identificarla.)

Prueba de optimalidad: Concluya que (0, 0) *no* es una solución óptima. (Las soluciones FEV adyacentes son mejores.)

Iteración 1: Muévase a una solución FEV *adyacente* mejor, (0, 6), y realice los siguientes tres pasos.

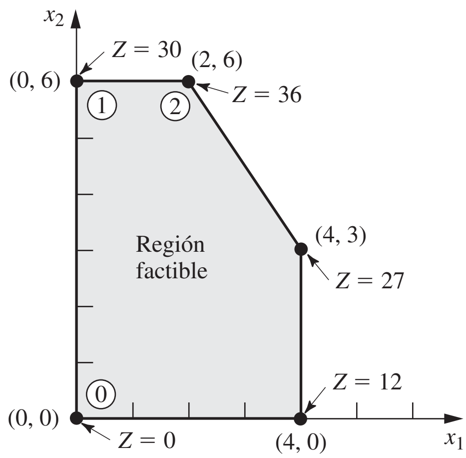
1. Entre las dos aristas de la región factible que salen de (0, 0), elija desplazarse a lo largo de la arista que aumenta el valor de x_2. (Con una función objetivo $Z = 3x_1 + 5x_2$, cuando aumenta el valor de x_2 el valor de Z crece más rápido que si aumentara el valor de x_1.)
2. Deténgase al llegar a la primera frontera de restricción: $2x_2 = 12$. [Si va más lejos en la dirección seleccionada en el paso 1, saldrá de la región factible; por ejemplo, cuando se desplaza hasta la segunda frontera de restricción en esa dirección se llega a (0, 9), que es una solución *no factible* en un vértice.]
3. Obtenga la intersección del nuevo conjunto de fronteras de restricción: (0, 6). (Las ecuaciones de estas fronteras de restricción, $x_1 = 0$ y $2x_2 = 12$, conducen de inmediato a esta solución.)

Prueba de optimalidad: Concluya que (0, 6) *no* es una solución óptima. (Existe una solución FEV adyacente que es mejor.)

Iteración 2: Muévase a una mejor solución FEV (2, 6), mediante la realización de los siguientes pasos.

1. En las dos aristas de la región factible que salen de (0, 6) elija moverse a lo largo de la que va hacia la derecha. (Moverse a lo largo de esta arista aumenta el valor de Z, mientras que ir para atrás hacia abajo del eje x_2 lo disminuye.)
2. Deténgase al encontrar la primera nueva frontera de restricción en esa dirección: $3x_1 + 2x_2 = 12$. (Si se mueve más lejos en la dirección elegida en el paso 1, saldrá de la región factible.)
3. Determine la intersección del nuevo conjunto de fronteras de restricción: (2, 6). (Las ecuaciones de estas fronteras de restricción, $3x_1 + 2x_2 = 18$ y $2x_2 = 12$, conducen de inmediato a esta solución.)

Prueba de optimalidad: Concluya que (2, 6) *es* una solución óptima y deténgase. (Ninguna de las soluciones FEV adyacentes es mejor.)



■ **FIGURA 4.2**
Esta gráfica muestra la secuencia de soluciones FEV (⓪, ①, ②) examinadas por el método símplex para el problema de la Wyndor Glass Co. La solución óptima (2, 6) se encuentra después de examinar tres soluciones.

En la figura 4.2 se muestra la secuencia de solución FEV que se examinó, donde cada número dentro de un círculo identifica la iteración que obtuvo esa solución. (Vea la sección de Ejemplos resueltos en al sitio en internet del libro donde encontrará **otro ejemplo** que muestra la forma en la que el método símplex itera a través de una secuencia de soluciones FEV antes de llegar a la solución óptima.)

A continuación se analizarán los seis conceptos clave para solución con el método símplex que proporcionan el razonamiento que lleva a los pasos anteriores. (Considere que estos conceptos también se aplican para resolver problemas con más de dos variables de decisión para los que no se dispone de una gráfica parecida a la que se muestra en la figura 4.2 como ayuda para encontrar una solución óptima.)

Conceptos clave de solución

El primer concepto de solución se basa de manera directa en la relación entre las soluciones óptimas y las soluciones factibles en los vértices explicada al final de la sección 3.2.

> **Concepto de solución 1:** El método símplex analiza sólo las soluciones FEV. Para cualquier problema con al menos una solución óptima, la ubicación de una de ellas sólo requiere encontrar una mejor solución FEV.[3]

Debido a que, por lo general, el número de soluciones factibles es infinito, la reducción del número de soluciones que deben examinarse a un pequeño número finito (sólo tres en la figura 4.2) es una simplificación enorme.

El siguiente concepto de solución define el flujo del método símplex.

> **Concepto de solución 2:** El método símplex es un *algoritmo iterativo* (procedimiento de solución sistemático que repite una serie fija de pasos, llamada *iteración*, hasta que se obtiene el resultado deseado) con la siguiente estructura.

Inicialización: Preparación para comenzar las iteraciones, que incluye encontrar una solución FEV inicial.

↓

→ **Prueba de optimalidad:** ¿Es óptima la solución FEV actual?

No ↓ Sí → Termina.

Iteración: Realizar una iteración para encontrar una mejor solución FEV.

[3] La única restricción es que el problema debe tener soluciones factibles en los vértices, lo cual se asegura si la región factible es acotada.

Observe que cuando se resolvió el ejemplo se siguió este diagrama de flujo de dos iteraciones hasta que se encontró una solución óptima.

Ahora se analizará la forma de comenzar.

Concepto de solución 3: Siempre que es posible, en el paso inicial del método símplex se elige el *origen* (todas las variables de decisión iguales a cero) como la solución FEV inicial. Cuando se tienen demasiadas variables de decisión para encontrar una solución FEV inicial en una gráfica, esta elección elimina la necesidad de usar procedimientos algebraicos para obtenerla.

Casi siempre es posible seleccionar el origen cuando todas las variables de decisión tienen restricciones de no negatividad, porque la intersección de estas fronteras de restricción lleva al origen como una solución en un vértice. En consecuencia, esta solución es una solución FEV *a menos* que sea *no factible* porque viole una o más restricciones funcionales. Si es no factible, se necesitan los procedimientos especiales descritos en la sección 4.6 para encontrar la solución FEV inicial.

El siguiente concepto de solución se refiere a la selección de una mejor solución FEV en cada iteración.

Concepto de solución 4: Dada una solución FEV, en términos de cómputo, es más rápido reunir información sobre sus soluciones FEV *adyacentes* que sobre otras soluciones FEV. Por lo tanto, cada vez que el método símplex realiza una iteración para moverse de la solución FEV actual a una mejor, *siempre* escoge una solución FEV *adyacente* a la actual. No considera otras soluciones FEV. En consecuencia, toda la trayectoria que sigue hasta alcanzar una solución óptima es a lo largo de las *aristas* de la región factible.

A continuación se verá cuál solución FEV adyacente se debe seleccionar en cada iteración.

Concepto de solución 5: Después de identificar la solución FEV actual, el método símplex examina cada una de las aristas de la región factible que salen de esta solución. Estas aristas conducen a una solución FEV *adyacente* en el otro punto extremo, pero el método símplex ni siquiera se toma la molestia de obtener la solución FEV adyacente. Sólo identifica la *tasa de mejoramiento de* Z que se obtendría al moverse por esa arista. Entre las aristas con una tasa *positiva* de mejoramiento de Z, selecciona moverse por aquella con la tasa *más grande* de mejoramiento de Z. La iteración termina cuando se obtiene primero la solución FEV al final de esta arista y después se reetiqueta esta solución FEV adyacente como la solución FEV *actual* para pasar a la prueba de optimalidad y (si es necesario) a la siguiente iteración.

En la primera iteración del ejemplo, al moverse de (0, 0) a lo largo de la arista sobre el eje x_1 se obtendría una tasa de mejoramiento de Z de 3 (Z crece 3 por cada incremento de una unidad de x_1), mientras que al moverse por la arista sobre el eje x_2 se tendría una tasa de mejoramiento de Z de 5 (Z aumenta 5 por cada incremento de una unidad de x_2), de manera que se toma la decisión de desplazarse a lo largo de esta última arista. En la segunda iteración, la única arista que sale de (0, 6) que daría una tasa de mejoramiento de Z *positiva* es la que conduce a (2, 6), por lo que se toma la decisión de moverse a lo largo de ella.

El concepto de solución final aclara la manera en que se realiza la prueba de optimalidad en forma eficiente.

Concepto de solución 6: El concepto de solución 5 describe la manera en que el método símplex examina cada arista de la región factible que sale de la solución FEV actual. A través de este examen de una arista es sencillo identificar la tasa de mejoramiento de Z que se obtendría al moverse por ella hasta la solución FEV adyacente en el otro extremo. Una tasa *positiva* de mejoramiento de Z implica que la solución FEV adyacente es *mejor* que la actual, mientras que una tasa *negativa* de mejoramiento de Z indica que la solución FEV adyacente es *peor*. Por lo tanto, la prueba de optimalidad consiste sólo en verificar si alguna de las aristas conduce a una tasa *positiva* de mejoramiento de Z. Si *ninguna* lo hace, la solución FEV actual es óptima.

En el ejemplo, al moverse por *cualquiera* de las dos aristas desde (2, 6), el valor de Z disminuye. Como se desea maximizar Z, este hecho lleva de inmediato a la conclusión de que (2, 6) es óptima.

4.2 PREPARACIÓN PARA EL MÉTODO SÍMPLEX

En la sección 4.1 se hizo hincapié en los conceptos geométricos fundamentales del método símplex. Sin embargo, lo común es que este algoritmo se trabaje en una computadora que sólo puede seguir instrucciones algebraicas. Por lo tanto, es necesario transformar el procedimiento geométrico conceptual que se acaba de describir en un procedimiento algebraico que se pueda usar. En esta sección se introducirá el *lenguaje algebraico* del método símplex y se relacionará con los conceptos de la sección anterior.

El procedimiento algebraico se basa en la solución de sistemas de ecuaciones. Por lo tanto, el primer paso para preparar el método símplex es convertir las *restricciones funcionales de desigualdad* en *restricciones de igualdad equivalentes*. (Las restricciones de no negatividad se dejan como desigualdades porque se manejan por separado.) Esta conversión se logra mediante la introducción de **variables de holgura**. Para ejemplificar, considere la primera restricción funcional del problema de la Wyndor Glass Co. de la sección 3.1

$$x_1 \leq 4.$$

La variable de holgura de esta restricción se define como

$$x_3 = 4 - x_1,$$

que es la holgura que queda en el lado izquierdo de la desigualdad. Entonces,

$$x_1 + x_3 = 4.$$

Dada esta ecuación, $x_1 \leq 4$ se cumple si y sólo si $4 - x_1 = x_3 \geq 0$. En consecuencia, la restricción original $x_1 \leq 4$ es por completo *equivalente* al par de restricciones

$$x_1 + x_3 = 4 \quad \text{y} \quad x_3 \geq 0.$$

Al introducir variables de holgura en las otras restricciones funcionales, el modelo de programación lineal original de este ejemplo (que se muestra debajo a la izquierda) se puede sustituir por el modelo equivalente (llamado *forma aumentada* del modelo) que se encuentra a la derecha:

Forma original del modelo

Maximizar $Z = 3x_1 + 5x_2,$

sujeta a

$$x_1 \leq 4$$
$$2x_2 \leq 12$$
$$3x_1 + 2x_2 \leq 18$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

Forma aumentada del modelo[4]

Maximizar $Z = 3x_1 + 5x_2,$

sujeta a

1) $x_1 + x_3 = 4$

2) $2x_2 + x_4 = 12$

3) $3x_1 + 2x_2 + x_5 = 18$

y

$$x_j \geq 0, \quad \text{para } j = 1, 2, 3, 4, 5.$$

Aun cuando ambas formas del modelo representan exactamente el mismo problema, la nueva forma es mucho más conveniente para la manipulación algebraica y la identificación de las soluciones FEV. Se le da el nombre de **forma aumentada** del problema, porque la forma original se *aumentó* con algunas variables suplementarias necesarias para aplicar el método símplex.

Si una variable de holgura es igual a 0 en la solución actual, entonces esta solución se encuentra sobre la frontera de restricción de la restricción funcional correspondiente. Un valor mayor que 0 significa que la solución está en el lado *factible* de la frontera de restricción, mientras que un valor menor que 0 señala que está en el lado *no factible* de esta frontera. La demostración de estas propiedades proporcionada por el **ejemplo de demostración** en el OR Tutor tiene el nombre de *Interpretation of the Slack Variables*.

[4] Las variables de holgura no se muestran en la función objetivo porque sus coeficientes son iguales a cero.

La terminología que se utiliza en la sección 4.1 (soluciones en los vértices, etc.) se aplica a la forma original del problema. A continuación se introducirá la terminología correspondiente a la forma aumentada.

Una **solución aumentada** es una solución de las variables originales (las *variables de decisión*) que se aumentó con los valores correspondientes de las *variables de holgura*.

Por ejemplo, si se aumenta la solución (3, 2) en el ejemplo, logra obtenerse la solución aumentada (3, 2, 1, 8, 5) debido a que los valores correspondientes de las variables de holgura son $x_3 = 1$, $x_4 = 8$ y $x_5 = 5$.

Una **solución básica** es una solución en un vértice *aumentada*.

Para ilustrar lo que decimos, considere la solución no factible del vértice (4, 6) de la figura 4.1. Al aumentarla con los valores que se obtuvieron para las variables de holgura $x_3 = 0$, $x_4 = 0$ y $x_5 = -6$ se obtiene la solución básica correspondiente (4, 6, 0, 0, −6).

El hecho de que las soluciones en los vértices (y por ende las soluciones básicas) puedan ser o no factibles implica la siguiente definición:

Una **solución básica factible (BF)** es una solución FEV *aumentada*.

Así, la solución FEV (0, 6) del ejemplo es equivalente a la solución BF (0, 6, 4, 0, 6) del problema en la forma aumentada.

La única diferencia entre las soluciones básicas y las soluciones en un vértice (o entre las soluciones BF y soluciones FEV) es el hecho de que están incluidos los valores de las variables de holgura. Dada cualquier solución básica, la solución en el vértice correspondiente se obtiene con sólo quitar las variables de holgura. En consecuencia, las relaciones geométricas y algebraicas entre estas dos soluciones son muy estrechas, como se verá en la sección 5.1.

Debido a que los términos *solución básica* y *solución básica factible* son partes muy importantes del vocabulario normal de programación lineal, es necesario aclarar sus propiedades algebraicas. En el caso de la forma aumentada del ejemplo observe que el sistema de restricciones funcionales tiene 5 variables y 3 ecuaciones, esto es,

Número de variables − número de ecuaciones = 5 − 3 = 2.

Este hecho proporciona 2 *grados de libertad* cuando se debe resolver el sistema puesto que se pueden elegir dos variables cualesquiera e igualarlas a cualquier valor arbitrario para resolver las tres ecuaciones en términos de las tres variables restantes.[5] El método símplex usa cero para este valor arbitrario. Así, dos de las variables (llamadas *variables no básicas*) se igualan a cero y, entonces, la solución simultánea de las tres ecuaciones de las otras tres variables (llamadas *variables básicas*) es una *solución básica*. Estas propiedades se describen en las siguientes definiciones generales.

Una **solución básica** tiene las siguientes propiedades:

1. Cada variable se designa ya sea como variable básica o como variable no básica.
2. El *número de variables básicas* es igual al número de restricciones funcionales (ahora ecuaciones). Por lo tanto, el *número de variables no básicas* es igual al número total de variables *menos* el número de restricciones funcionales.
3. Las **variables no básicas** se igualan a cero.
4. Los valores de las **variables básicas** se obtienen como la solución simultánea del sistema de ecuaciones (restricciones funcionales en la forma aumentada). (Con frecuencia se hace referencia al conjunto de variables básicas como **la base**.)
5. Si las variables básicas satisfacen las *restricciones de no negatividad*, la solución básica es una **solución BF**.

[5] Este método de determinar el número de grados de libertad de un sistema de ecuaciones es válido *siempre y cuando* el sistema no incluya ecuaciones redundantes. Esta condición siempre se cumple para el sistema de ecuaciones formado por las restricciones funcionales en la forma aumentada de un modelo de programación lineal.

Con el fin de ilustrar estas definiciones se considerará de nuevo la solución BF (0, 6, 4, 0, 6). Esta solución se obtuvo al aumentar la solución FEV (0, 6). Sin embargo, otra manera de obtenerla es elegir x_1 y x_4 como variables no básicas, es decir, como las dos variables que se igualan a cero. Las tres ecuaciones respectivas llevan a $x_3 = 4$, $x_2 = 6$ y $x_5 = 6$ como la solución de las tres variables básicas, según se muestra en seguida (las variables básicas aparecen en **negritas**):

$$
\begin{array}{llr}
 & & x_1 = 0 \text{ y } x_4 = 0 \text{ entonces} \\
(1) & x_1 + \mathbf{x_3} = 4 & \mathbf{x_3} = 4 \\
(2) & 2\mathbf{x_2} + x_4 = 12 & \mathbf{x_2} = 6 \\
(3) & 3x_1 + 2\mathbf{x_2} + \mathbf{x_5} = 18 & \mathbf{x_5} = 6
\end{array}
$$

Como estas tres variables básicas son no negativas, esta *solución básica* (0, 6, 4, 0, 6) sin duda es una *solución BF*. En la sección Worked Examples del sitio en internet del libro se incluye **otro ejemplo** de la relación entre las soluciones FEV y BF.

Como sólo ciertos pares de soluciones FEV son *adyacentes*, a los pares correspondientes de soluciones BF se les denomina adyacentes. Un modo simple de reconocer si dos soluciones BF son adyacentes es el siguiente.

> Dos soluciones BF son **adyacentes** si *todas menos una* de sus *variables no básicas* son las mismas. Esto implica que *todas menos una* de sus *variables básicas* son también las mismas, aunque tal vez con valores numéricos diferentes.

En consecuencia, al trasladarse de la solución BF actual a una adyacente se cambia una variable no básica a básica y viceversa para otra variable (y después se ajustan los valores de las variables básicas para que satisfagan el sistema de ecuaciones).

Para dar un ejemplo de *soluciones básicas factibles adyacentes*, considere un par de soluciones factibles en vértices adyacentes en la figura 4.1: (0, 0) y (0, 6). Sus soluciones aumentadas (0, 0, 4, 12, 18) y (0, 6, 4, 0, 6) son, de manera automática, soluciones BF adyacentes. Sin embargo, no es necesario ver la figura 4.1 para llegar a esta conclusión. Otra forma de verlo es observar que sus variables no básicas (x_1, x_2) y (x_1, x_4), son las mismas excepto en una, x_4 sustituye a x_2. En consecuencia, trasladarse de (0, 0, 4, 12, 18) a (0, 6, 4, 0, 6) implica cambiar a x_2 de variable no básica a básica y lo contrario para x_4.

Cuando se trabaja con el problema en la forma aumentada conviene tomar en cuenta y manipular la ecuación de la función objetivo al mismo tiempo que las nuevas ecuaciones de las restricciones. Antes de comenzar con el método símplex es necesario escribir el problema una vez más en una forma equivalente:

Maximizar Z,

sujeta a

$$
\begin{array}{ll}
(0) & Z - 3x_1 - 5x_2 = 0 \\
(1) & x_1 + x_3 = 4 \\
(2) & 2x_2 + x_4 = 12 \\
(3) & 3x_1 + 2x_2 + x_5 = 18
\end{array}
$$

y

$x_j \geq 0$, para $j = 1, 2, \ldots, 5$.

Es justo como si la ecuación (0) fuera una de las restricciones originales; pero como ya se encuentra en forma de igualdad, no necesita variable de holgura. Al mismo tiempo que se agregó una ecuación, también se añadió una variable (Z) al sistema de ecuaciones. Por lo tanto, al usar las ecuaciones (1) a (3) para obtener una solución básica como se describió, se usa la ecuación (0) para encontrar al mismo tiempo el valor de Z.

Por casualidad, el modelo del problema de la Wyndor Glass Co. se ajusta a *nuestra forma estándar*, y todas sus restricciones funcionales tienen valores no negativos b_i en el lado derecho. Si el caso fuera diferente, sería necesario hacer ajustes adicionales en este punto antes de aplicar el método símplex. Estos detalles se abordarán en la sección 4.6 para ahora dedicar toda la atención a este método.

4.3 ÁLGEBRA DEL MÉTODO SÍMPLEX

Con propósitos ilustrativos se seguirá con el ejemplo modelo de la sección 3.1, como se escribió al final de la sección 4.2. Para comenzar a relacionar los conceptos geométricos con los algebraicos del método símplex, en la tabla 4.2 se describirán, en paralelo, los pasos que siguen el enfoque geométrico y el algebraico para resolver este ejemplo. Como el punto de vista geométrico (que se presentó por primera vez en la sección 4.1) se basa en la *forma original* del modelo (sin variables de holgura), es necesario consultar de nuevo la figura 4.1 para tener una imagen mental de lo descrito en la segunda columna de la tabla. Examine la *forma aumentada* del modelo que se presenta al final de la sección 4.2 cuando se analice la tercera columna de la tabla.

A continuación se darán los detalles de cada paso en la tercera columna de la tabla 4.2.

TABLA 4.2 Interpretaciones geométrica y algebraica de los pasos del método símplex para resolver el ejemplo de la Wyndor Glass Co.

Secuencia del método	Interpretación geométrica	Interpretación algebraica
Paso inicial	Se selecciona (0, 0) como la solución FEV inicial.	Se selecciona x_1 y x_2 como variables no básicas (= 0) para la solución BF inicial: (0, 0, 4, 12, 18).
Prueba de optimalidad	No es óptima porque al moverse por cualquier arista que parte desde (0, 0), Z crece.	No es óptima porque cuando aumenta el valor de cualquier variable no básica (x_1 o x_2), el valor de Z aumenta.
Iteración 1		
Paso 1	Se mueve por la arista sobre el eje x_2.	Se aumenta el valor de x_2 y se ajustan los valores de las otras variables para que satisfagan el sistema de ecuaciones.
Paso 2	Se detiene cuando encuentra la primera frontera de restricción ($2x_2 = 12$).	Se detiene cuando el valor de la primera variable básica (x_3, x_4 o x_5) llega a cero (x_4).
Paso 3	Se encuentra la intersección del nuevo par de fronteras de restricción: (0, 6) es la nueva solución FEV.	Con x_2 como variable básica y x_4 no básica, se resuelve el sistema de ecuaciones: (0, 6, 4, 0, 6) es la nueva solución BF.
Prueba de optimalidad	No es óptima porque al moverse por la arista que parte desde (0, 6) hacia la derecha, Z crece.	No es óptima porque cuando aumenta el valor de una variable no básica (x_1), el valor de Z aumenta.
Iteración 2		
Paso 1	Se mueve por la arista que va hacia la derecha.	Se aumenta el valor x_1 y se ajustan los de las demás variables para que satisfagan el sistema de ecuaciones.
Paso 2	Se detiene cuando encuentra la primera frontera de restricción ($3x_1 + 2x_2 = 18$).	Se detiene cuando la primera variable básica (x_2, x_3 o x_5) llega a cero (x_5).
Paso 3	Se encuentra la intersección del nuevo par de fronteras de restricción: (2, 6) es la nueva solución FEV.	Con x_1 como variable básica y x_5 no básica, se resuelve el sistema de ecuaciones: (2, 6, 2, 0, 0) es la nueva solución BF.
Prueba de optimalidad	(2, 6) es óptima porque al moverse por cualquier arista que sale de (2, 6), Z decrece.	(2, 6, 2, 0, 0) es óptima porque cuando aumenta el valor de cualquier variable no básica (x_4 o x_5) el valor de Z disminuye.

Recuadro de aplicación

La compañía **Samsung Electronics Corp., Ltd. (SEC)** es líder en la fabricación de memorias de acceso aleatorio estático y dinámico y de otros circuitos integrados digitales avanzados. En sus instalaciones en Kiheung, Corea del Sur (probablemente sea la fábrica más grande de semiconductores del mundo), se producen más de 300 000 obleas de silicio al mes, y da trabajo a más de 10 000 empleados.

La *duración del ciclo* es el término que se usa en la industria para definir el tiempo que transcurre desde que se ingresa un lote de obleas de silicio en blanco al proceso de fabricación hasta que se terminan de fabricar los dispositivos en ellas. La reducción de la duración del ciclo es un objetivo constante ya que así se reducen costos y se pueden ofrecer tiempos de entrega más cortos a los clientes potenciales, lo que significa un aspecto clave para mantener o incrementar la participación de mercado en una industria extremadamente competitiva.

Existen tres factores que, en particular, presentan grandes retos cuando se quiere reducir la duración de los ciclos. Uno de ellos estriba en el hecho de que la mezcla de productos cambia de manera continua. Otro es que, a menudo, la compañía debe llevar a cabo cambios significativos en el programa de fabricación dentro de la duración del ciclo objetivo a medida que revisa los pronósticos de la demanda del cliente. El tercero es que las máquinas de un tipo genérico son heterogéneas de tal forma que sólo un reducido número de ellas son aptas para realizar cada etapa del proceso de fabricación del dispositivo.

Un equipo especializado en investigación de operaciones desarrolló *un modelo de programación lineal de gran tamaño con decenas de miles de variables de decisión y restricciones funcionales* para poder lidiar con estos retos. La función objetivo involucraba minimizar las órdenes de compra y el inventario de los bienes terminados. A pesar del enorme tamaño de este modelo, se resolvía en cuestión de minutos siempre que se necesitaba mediante el uso de una compleja implementación del método símplex (y ciertas técnicas relacionadas) en el software de optimización CPLEX. (Este software se estudiará más adelante en la sección 4.8.)

La implantación continua de este modelo hizo posible que la compañía redujera los tiempos de fabricación de memorias dinámicas de acceso aleatorio de más de 80 días a menos de 30. Esta mejora significativa y la consecuente reducción de los costos de fabricación y los precios de venta hicieron posible que Samsung obtuviera una *ganancia adicional de* **200 millones de dólares** *en ventas anuales.*

Fuente: R.C. Leachman, J. Kang y Y. Lin, "SLIM: Short Cycle Time and Low Inventory in Manufacturing at Samsung Electronics", en *Interfaces*, **32**(1): 61-77, enero-febrero de 2002. (En nuestro sitio en internet www.mhhe.com/hillier se puede encontrar una liga hacia este artículo.)

Paso inicial

La elección de x_1 y x_2 como las variables *no básicas* (las variables que se igualan a cero) para la solución BF inicial se basa en el concepto de solución 3 de la sección 4.1. Esta elección elimina el trabajo que se requiere para despejar las *variables básicas* (x_3, x_4, x_5) del siguiente sistema de ecuaciones (donde las variables básicas se muestran en **negritas**):

$$\begin{array}{llr} & & x_1 = 0 \text{ y } x_2 = 0 \text{ entonces} \\ (1) & x_1 + \boldsymbol{x_3} = 4 & \boldsymbol{x_3} = 4 \\ (2) & 2x_2 + \boldsymbol{x_4} = 12 & \boldsymbol{x_4} = 12 \\ (3) & 3x_1 + 2x_2 + \boldsymbol{x_5} = 18 & \boldsymbol{x_5} = 18 \end{array}$$

En consecuencia, la **solución BF inicial** es (0, 0, 4, 12, 18).

Observe que esta solución se puede leer de inmediato porque cada ecuación tiene sólo una variable básica con coeficiente 1, la cual no aparece en ninguna otra ecuación. Pronto se verá que cuando cambia el conjunto de variables básicas, el método símplex recurre a un procedimiento algebraico (eliminación gaussiana) para convertir las ecuaciones en la misma forma conveniente para leer las soluciones BF subsecuentes. Esta forma se llama **forma apropiada de eliminación de Gauss**.

Prueba de optimalidad

La función objetivo es

$$Z = 3x_1 + 5x_2,$$

de manera que $Z = 0$ en la solución BF inicial. Debido a que ninguna variable básica (x_3, x_4, x_5) tiene coeficiente *distinto de cero* en esta función objetivo, el coeficiente de las variables no básicas (x_1, x_2) proporciona la tasa de mejoramiento de Z si se aumentara el valor de esa variable a más de cero (mientras que los valores de las variables básicas se ajustan

para que satisfagan el sistema de ecuaciones).[6] Estas tasas de mejoramiento (3 y 5) son *positivas*. Por lo tanto, según el concepto de solución 6 de la sección 4.1, se concluye que (0, 0, 4, 12, 18) no es óptima.

En el caso de cada una de las soluciones BF que se examinan después de varias iteraciones, al menos una variable básica tiene coeficiente diferente de cero en la función objetivo. Por lo tanto, la prueba de optimalidad usará la nueva ecuación (0) para reescribir la función objetivo en términos de las variables no básicas, como se verá más adelante.

Determinación de la dirección de movimiento (paso 1 de una iteración)

Cuando aumenta el valor de una variable no básica (y se ajustan los valores de las variables básicas para que satisfagan el sistema de ecuaciones) se realiza una acción equivalente a moverse a lo largo de una arista que sale de la solución FEV actual. Según los conceptos de solución 4 y 5 de la sección 4.1, la elección de cuál variable no básica debe aumentar su valor se hace como sigue:

$$
\begin{array}{lll}
 & Z = 3x_1 + 5x_2 & \\
\text{¿Aumenta } x_1? & \text{Tasa de mejoramiento} & Z = 3. \\
\text{¿Aumenta } x_2? & \text{Tasa de mejoramiento} & Z = 5.
\end{array}
$$

$5 > 3$, de manera que se elije x_2 para aumentar su valor

Como se indica en seguida, x_2 se llama *variable básica entrante* de la iteración 1.

> En cada iteración del método símplex el propósito del paso 1 es elegir una *variable no básica* para que aumente su valor (y se ajustan los valores de las variables básicas para que satisfagan el sistema de ecuaciones). Si aumenta el valor de esta variable no básica se convertirá en *variable básica* de la siguiente solución BF. Por lo tanto, esta variable se llama **variable básica entrante** de la iteración actual (porque es la que entrará a la base).

Determinación de dónde detenerse (paso 2 de una iteración)

El paso 2 contesta la pregunta de cuánto aumentar el valor de la variable básica entrante x_2 antes de detenerse. Si aumenta la variable x_2 crece el valor de Z, el cual se desea incrementar lo más posible sin salirse de la región factible. El requerimiento de satisfacer las restricciones funcionales de la forma aumentada (que se muestra en seguida) significa que al aumentar x_2 (al mismo tiempo que se mantiene la variable no básica $x_1 = 0$) cambia el valor de algunas variables básicas, como se observa en el lado derecho.

$$
\begin{array}{llll}
 & & x_1 = 0, & \text{por lo tanto} \\
(1) & x_1 + \boldsymbol{x_3} = 4 & \boldsymbol{x_3} = 4 & \\
(2) & 2x_2 + \boldsymbol{x_4} = 12 & \boldsymbol{x_4} = 12 - 2x_2 & \\
(3) & 3x_1 + 2x_2 + \boldsymbol{x_5} = 18 & \boldsymbol{x_5} = 18 - 2x_2. &
\end{array}
$$

El otro requisito para lograr la factibilidad es que todas las variables sean *no negativas*. Las variables no básicas (entre ellas la variable entrante) son no negativas, pero es necesario verificar cuánto puede crecer x_2 sin violar las restricciones de no negatividad de las variables básicas.

$$
\begin{array}{ll}
\boldsymbol{x_3} = 4 \geq 0 & \Rightarrow \text{no hay cota superior sobre } x_2. \\
\boldsymbol{x_4} = 12 - 2x_2 \geq 0 \Rightarrow x_2 \leq \dfrac{12}{2} = 6 & \leftarrow \text{mínimo.} \\
\boldsymbol{x_5} = 18 - 2x_2 \geq 0 \Rightarrow x_2 \leq \dfrac{18}{2} = 9. &
\end{array}
$$

En consecuencia, x_2 puede crecer exactamente hasta 6, punto en el que $\boldsymbol{x_4}$ ha llegado a 0. Si se aumentara x_2 a más de 6 se provocaría que $\boldsymbol{x_4}$ adquiriese signo negativo, lo que violaría la factibilidad.

[6] Observe que esta interpretación de los coeficientes de x_j se basa en que estas variables se encuentren en el lado derecho, $Z = 3x_1 + 5x_2$. Cuando se pasan al lado izquierdo de la ecuación (0), $Z - 3x_1 - 5x_2 = 0$, los coeficientes distintos de cero cambian su signo.

Estos cálculos reciben el nombre de **prueba del cociente mínimo**. El objetivo de la prueba es determinar qué variable básica llega a cero primero cuando crece la variable entrante. De inmediato se puede descartar la variable básica en las ecuaciones donde el coeficiente de la variable básica entrante es cero o negativo, puesto que estas variables no decrecen si la variable básica entrante aumenta. [Esto ocurrió con x_3 en la ecuación (1) del ejemplo.] Sin embargo, por cada ecuación donde el coeficiente de la variable básica entrante es *estrictamente positivo* (> 0), esta prueba calcula el *cociente* del lado derecho entre el coeficiente de la variable básica entrante. La variable básica de la ecuación con el *cociente mínimo* es la que llega a cero primero cuando crece la variable básica entrante.

> En cualquier iteración del método símplex, el paso 2 utiliza la prueba del *cociente mínimo* para determinar cuál variable básica llega primero a cero cuando aumenta el valor de la variable básica entrante. Al disminuir hasta cero el valor de esta variable básica se convierte en *variable no básica* de la siguiente solución BF. Por lo tanto, esta variable se llama **variable básica saliente** de la iteración actual (puesto que es la que deja la base).

De esta manera, $\boldsymbol{x_4}$ es la variable básica que sale de la iteración 1 del ejemplo.

Resolución de una nueva solución BF (paso 3 de una iteración)

Cuando aumenta $x_2 = 0$ hasta $x_2 = 6$ el procedimiento se mueve de la solución BF *inicial* que se muestra a la izquierda hacia la *nueva* solución BF de la derecha.

	Solución BF inicial	Nueva solución BF
Variables no basicas:	$x_1 = 0, \quad x_2 = 0$	$x_1 = 0, \quad x_4 = 0$
Variables básicas:	$x_3 = 4, \quad x_4 = 12, \quad x_5 = 18$	$x_3 = ?, \quad x_2 = 6, \quad x_5 = ?$

El propósito del paso 3 es convertir el sistema de ecuaciones en una forma más conveniente (la forma adecuada de eliminación gaussiana) para llevar a cabo la prueba de optimalidad y la siguiente iteración (si es necesario) con la nueva solución BF. En el proceso, esta forma también identificará los valores de x_3 y x_5 de la nueva solución.

Se escribirá de nuevo el sistema de ecuaciones completo, con las *nuevas* variables básicas en **negritas** (Z es la variable básica en la ecuación de la función objetivo):

$$\begin{array}{llr}
(0) & \boldsymbol{Z} - 3x_1 - 5\boldsymbol{x_2} & = 0 \\
(1) & x_1 + \boldsymbol{x_3} & = 4 \\
(2) & 2\boldsymbol{x_2} + x_4 & = 12 \\
(3) & 3x_1 + 2\boldsymbol{x_2} + \boldsymbol{x_5} & = 18.
\end{array}$$

Entonces, x_2 sustituyó a x_4 como la variable básica en la ecuación (2). Para despejar Z, x_2, x_3 y x_5 de este sistema de ecuaciones es necesario realizar algunas **operaciones algebraicas elementales** para reproducir el patrón actual de coeficientes de x_4 (0, 0, 1, 0) como los nuevos coeficientes de x_2. Se pueden realizar cualquiera de los dos tipos de operaciones algebraicas elementales:

1. Multiplicar (o dividir) una ecuación por una constante distinta de cero.
2. Sumar (o restar) un múltiplo de una ecuación a (o de) otra ecuación.

Para preparar la ejecución de estas operaciones, observe que los respectivos coeficientes de x_2 en el sistema de ecuaciones anterior son -5, 0, 2 y 2, y se intenta convertirlos en 0, 0, 1 y 0, respectivamente. Para convertir el coeficiente 2 de la ecuación (2) en un 1, se usa el primer tipo de operación algebraica elemental y se divide esta ecuación entre 2 para obtener

$$(2) \qquad \boldsymbol{x_2} + \frac{1}{2}x_4 = 6.$$

Para convertir los coeficientes -5 y 2 en ceros, es necesario usar un segundo tipo de operación elemental. En particular, se suma a la ecuación (0) esta nueva ecuación (2) multiplicada por 5, y

se resta de la ecuación (3) esta nueva ecuación multiplicada por 2. El nuevo sistema de ecuaciones que resulta es

$$\begin{array}{llcl}
(0) & Z - 3x_1 \qquad + \frac{5}{2}x_4 & = & 30 \\
(1) & \quad x_1 + x_3 & = & 4 \\
(2) & \quad x_2 \qquad + \frac{1}{2}x_4 & = & 6 \\
(3) & \quad 3x_1 \qquad - x_4 + x_5 & = & 6.
\end{array}$$

Como $x_1 = 0$ y $x_4 = 0$, las ecuaciones de esta forma llevan de inmediato a la nueva solución BF, $(x_1, x_2, x_3, x_4, x_5) = (0, 6, 4, 0, 6)$, de donde se obtiene $Z = 30$.

Este procedimiento para obtener la solución simultánea de un sistema de ecuaciones lineales se llama *método de eliminación de Gauss-Jordan*, o, de manera más breve, **eliminación gaussiana**.[7] El concepto clave de este método es usar las operaciones algebraicas elementales para reducir el sistema de ecuaciones original a la forma apropiada de eliminación gaussiana, en donde cada variable básica se elimina de todas las ecuaciones menos de una (*su* ecuación), en la cual tiene coeficiente +1.

Prueba de optimalidad de la nueva solución básica factible

La ecuación (0) actual proporciona el valor de la función objetivo en términos nada más de las variables no básicas actuales,

$$Z = 30 + 3x_1 - \frac{5}{2}x_4.$$

Al aumentar el valor de cualquiera de estas variables no básicas (con el ajuste de los valores de las variables básicas para que cumplan con el sistema de ecuaciones) se realiza una acción equivalente a trasladarse a una de las dos soluciones BF *adyacentes*. Como x_1 tiene coeficiente *positivo*, cuando crece conduce a una solución BF adyacente que es mejor que la solución actual, por lo que ésta todavía no es óptima.

Iteración 2 y la solución óptima que resulta

Como $Z = 30 + 3x_1 - \frac{5}{2}x_4$, Z se puede aumentar si aumenta el valor de x_1, pero no el de x_4. Por lo tanto, el paso 1 elige a x_1 como la variable básica entrante.

En el paso 2, el sistema de ecuaciones actual lleva a las siguientes conclusiones sobre qué tanto se puede aumentar x_1 (con $x_4 = 0$):

$$x_3 = 4 - x_1 \geq 0 \quad \Rightarrow x_1 \leq \frac{4}{1} = 4.$$

$$x_2 = 6 \geq 0 \quad \Rightarrow \text{no hay cota superior sobre } x_1.$$

$$x_5 = 6 - 3x_1 \geq 0 \Rightarrow x_1 \leq \frac{6}{3} = 2 \quad \leftarrow \text{mínimo.}$$

Por lo tanto, la prueba del cociente mínimo indica que x_5 es la variable básica que sale.

En el paso 3 se sustituye x_1 por x_5 como variable básica, se realizan las operaciones algebraicas elementales en el sistema de ecuaciones actual para reproducir el patrón de coeficientes de x_5 (0, 0, 0, 1) como los nuevos coeficientes de x_1. De aquí se obtiene el siguiente nuevo sistema de ecuaciones:

$$\begin{array}{llcl}
(0) & Z \qquad + \frac{3}{2}x_4 + x_5 & = & 36 \\
(1) & \quad x_3 + \frac{1}{3}x_4 - \frac{1}{3}x_5 & = & 2 \\
(2) & \quad x_2 \quad + \frac{1}{2}x_4 & = & 6 \\
(3) & \quad x_1 \qquad - \frac{1}{3}x_4 + \frac{1}{3}x_5 & = & 2.
\end{array}$$

[7] En realidad existen algunas diferencias técnicas entre el método de eliminación de Gauss-Jordan y la eliminación gaussiana, pero no se hará esta distinción.

Por lo tanto, la siguiente solución BF es $(x_1, x_2, x_3, x_4, x_5) = (2, 6, 2, 0, 0)$, de donde se obtiene $Z = 36$. Para aplicar la *prueba de optimalidad* a esta nueva solución BF se usa la ecuación (0) actual para expresar Z sólo en términos de las variables no básicas actuales,

$$Z = 36 - \frac{3}{2}x_4 - x_5.$$

Cuando se incrementa x_4 o x_5 el valor de Z *disminuirá*, de manera que ninguna solución BF adyacente es tan buena como la actual. Entonces, según el concepto de solución 6 de la sección 4.1, la solución BF actual debe ser óptima.

En términos de la forma original del problema (sin variables de holgura), la solución óptima es $x_1 = 2, x_2 = 6$, lo que lleva a $Z = 3x_1 + 5x_2 = 36$.

Para ver **otro ejemplo** de aplicación del método símplex se recomienda que en este momento se vea, en el OR Tutor, la demostración llamada *Simplex Method—Algebraic Form* (método símplex: forma algebraica). Esta demostración despliega de manera simultánea el álgebra y la geometría del método símplex y va paso a paso en forma dinámica. Igual que en otros ejemplos de demostración que acompañan otras secciones del libro (incluso la siguiente), esta demostración computacional pone de relieve los conceptos que son difíciles de plasmar en el papel. Además, en la sección Worked Examples del sitio en internet del libro se incluye **otro ejemplo** de aplicación del método símplex.

Para ayudar más en el aprendizaje eficiente del método símplex, el tutorial IOR en su OR Courseware incluye un procedimiento llamado **Solve Interactively by the Simplex Method** (solución interactiva por el método símplex). Esta rutina realiza todos los cálculos mientras se toman las decisiones paso a paso, lo cual permite concentrarse en los conceptos en lugar de hacerlo en todos los números y cálculos. En consecuencia, podría ser recomendable usar esta rutina para resolver los problemas de esta sección. El software resulta útil en el inicio del empleo del método puesto que indica cuando se comete un error en la primera iteración de un problema.

Después de aprender el método símplex, el lector deseará simplemente aplicar una implantación computacional de éste para obtener soluciones óptimas inmediatas para los problemas de programación lineal. Por ello, se ha considerado conveniente también incluir en el IOR Tutorial un procedimiento automático llamado **Solve Automatically by the Simplex Method** (solución automática por el método símplex). Este procedimiento está diseñado para tratar sólo con problemas del tamaño de los que se incluyen en libros de texto y, además, cuenta con un procedimiento interactivo que incluye la comprobación de respuestas. En la sección 4.8 se describirán opciones de software para programación lineal más poderosas, que también se proporcionan en el sitio en internet del libro.

La siguiente sección incluye un resumen del método símplex en una forma tabular más conveniente.

4.4 EL MÉTODO SÍMPLEX EN FORMA TABULAR

La forma algebraica del método símplex que se presentó en la sección 4.3 puede ser la mejor para entender la lógica que fundamenta el algoritmo. Sin embargo, no es la más conveniente para realizar los cálculos necesarios. Cuando se tenga que resolver un problema a mano (o en forma interactiva con su tutorial IOR) se recomienda la *forma tabular* descrita en esta sección.[8]

La forma tabular del método símplex registra sólo la información esencial, a saber: 1) los coeficientes de las variables, 2) las constantes del lado derecho de las ecuaciones y 3) la variable básica que aparece en cada ecuación. Esta forma evita tener que escribir los símbolos de las variables en cada ecuación, pero es más importante el hecho de que permite hacer hincapié en los números que se usan en los cálculos aritméticos y registrarlos en forma muy compacta.

En la tabla 4.3 se hace una comparación entre el sistema de ecuaciones inicial del problema de la Wyndor Glass Co. en forma algebraica (en el lado izquierdo) y la forma tabular (a la derecha), donde la tabla de la derecha se conoce como *tabla símplex*. La variable básica que aparece en cada

[8] En la sección 5.2 se presenta una forma más adecuada para ejecutarse en la computadora.

TABLA 4.3 Tabla símplex inicial del problema de la Wyndor Glass Co.

a) Forma algebraica	*b*) Forma tabular									
	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	x_5	Lado derecho
(0) $\mathbf{Z} - 3x_1 - 5x_2 = 0$	Z	(0)	1	−3	−5	0	0	0	0
(1) $x_1 + \mathbf{x_3} = 4$	x_3	(1)	0	1	0	1	0	0	4
(2) $2x_2 + \mathbf{x_4} = 12$	x_4	(2)	0	0	2	0	1	0	12
(3) $3x_1 + 2x_2 + \mathbf{x_5} = 18$	x_5	(3)	0	3	2	0	0	1	18

ecuación se muestra en **negritas** en la columna de la izquierda y en la primera columna de la tabla símplex de la derecha. [Aunque sólo las variables x_j son básicas o no básicas, Z tiene el papel de la variable básica para la ecuación (0).] En forma automática se sabe que las variables que *no* se enumeran en esta columna de *variables básicas* (x_1, x_2) son *variables no básicas*. Después de establecer $x_1 = 0, x_2 = 0$, la columna denominada *lado derecho* proporciona la solución de las variables básicas, de manera que la solución BF inicial es $(x_1, x_2, x_3, x_4, x_5) = (0, 0, 4, 12, 18)$ con $Z = 0$.

> La *forma tabular* del método símplex utiliza una **tabla símplex** para desplegar de manera compacta el sistema de ecuaciones que conduce a la solución BF actual. De acuerdo con esta solución, cada variable de la columna de la izquierda es igual al número correspondiente en la columna de la derecha (y las variables que no aparecen son iguales a cero). Cuando se realiza la prueba de optimalidad o una iteración, los únicos números relevantes son los que están a la derecha de la columna Z.[9] El término **renglón** se refiere a una fila horizontal de números a la derecha de la columna Z (que incluyen el número del *lado derecho*), donde el renglón i corresponde a la ecuación (i).

En seguida se resume la forma tabular del método símplex y, al mismo tiempo, en forma breve, se describe su aplicación al problema de Wyndor Glass Co. No debe perderse de vista que la lógica es idéntica a la de la forma algebraica que se presentó en la sección anterior; sólo cambia la forma para desplegar tanto el sistema de ecuaciones actual como la iteración subsecuente (además de que ya no se tendrán que pasar las variables al lado derecho de una ecuación antes de llegar a las conclusiones de la prueba de optimalidad o en los pasos 1 y 2 de una iteración).

Resumen del método símplex (y la iteración 1 del ejemplo)

Paso inicial. Se introducen las variables de holgura. Se seleccionan las *variables de decisión* como las *variables no básicas iniciales* (es decir, iguales a cero) y las *variables de holgura* como las *variables básicas iniciales*. (Vea la sección 4.6 para hacer los ajustes necesarios al modelo si no se encuentra en nuestra forma estándar —maximización, sólo restricciones funcionales tipo $\leq$ y todas las restricciones de no negatividad— o si el valor de alguna b_i es negativo.)

En el ejemplo: Esta selección conduce a la tabla símplex inicial que se muestra en la tabla 4.3, por lo que la solución BF inicial es (0, 0, 4, 12, 18).

Prueba de optimalidad. La solución BF es óptima si y sólo si *todos* los coeficientes del renglón 0 son no negativos (≥ 0). Si es así, el proceso se detiene; de otra manera, sigue a una iteración para obtener la siguiente solución BF, que incluye cambiar una variable no básica a básica (paso 1) y viceversa (paso 2) y después despejar la nueva solución (paso 3).

En el ejemplo: Igual que $Z = 3x_1 + 5x_2$ indica que al aumentar x_1 o x_2 el valor de Z aumenta, de manera que la solución BF actual no es óptima, se llega a la misma conclusión a partir de la ecuación $Z - 3x_1 - 5x_2 = 0$. Los coeficientes -3 y -5 se muestran en el renglón 0 de la tabla 4.3.

Iteración. *Paso 1*: Se determina la *variable básica entrante* con la selección de la variable (que de modo automático es no básica) con el *coeficiente negativo* que tiene el mayor valor absoluto

[9] Por esta razón, se pueden suprimir la ecuación y las columnas correspondientes a Z con el fin de reducir el tamaño de la tabla símplex. Preferimos conservar estas columnas como recordatorio de que la tabla símplex está desplegando el sistema actual de ecuaciones y que Z es una de las variables de la ecuación (0).

TABLA 4.4 Aplicación de la prueba del cociente mínimo para determinar la primera variable básica saliente en el problema de la Wyndor Glass Co.

Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	x_5	Lado derecho	Cociente
Z	(0)	1	−3	−5	0	0	0	0	
x_3	(1)	0	1	0	1	0	0	4	
x_4	(2)	0	0	2	0	1	0	12	$\rightarrow \frac{12}{2} = 6 \leftarrow$ mínimo
x_5	(3)	0	3	2	0	0	1	18	$\rightarrow \frac{18}{2} = 9$

(es decir, el coeficiente "más negativo") de la ecuación (0). Se pone un recuadro alrededor de la columna abajo de este coeficiente y se le da el nombre de **columna pivote**.

En el ejemplo: El coeficiente más negativo es -5 para x_2 ($5 > 3$), de manera que x_2 debe convertirse en variable básica. (Este cambio se indica en la tabla 4.4 mediante el recuadro alrededor de la columna abajo de -5.)

Paso 2: Se determina la *variable básica que sale* con la prueba del cociente mínimo.

Prueba del cociente mínimo

1. Elija los coeficientes estrictamente positivos (> 0) de la columna pivote.
2. Divida el elemento del *lado derecho* del mismo renglón entre dicho coeficiente.
3. Identifique el renglón que tiene el *menor* de estos cocientes.
4. La variable básica de ese renglón es la variable básica que sale; sustitúyala con la variable básica entrante en la columna de la variable básica de la siguiente tabla.

Ponga un recuadro en este renglón que se llama **renglón pivote**. El número que se encuentra en *ambos* recuadros se llama **número pivote**.

En el ejemplo: Los cálculos de la prueba del cociente mínimo se muestran a la derecha de la tabla 4.4. El renglón 2 es el renglón pivote (vea el recuadro alrededor de ese renglón en el primer cuadro símplex de la tabla 4.5), mientras que x_4 es la variable básica que sale. En la siguiente tabla símplex (vea la tabla 4.5) x_2 sustituye a x_4 como la variable básica del renglón 2.

Paso 3: Se despeja la *nueva solución BF* mediante **operaciones elementales con renglones** (multiplicación o división de un renglón por una constante diferente de cero; suma o resta de un múltiplo de un renglón con otro) para construir una nueva tabla símplex en la forma apropiada de eliminación gaussiana, abajo de la tabla actual, y después se regresa a la prueba de optimalidad. Las operaciones elementales con renglones que deben realizarse son:

TABLA 4.5 Tabla símplex del problema de la Wyndor Glass Co. despúes de dividir el primer renglón pivote entre el primer número pivote

Iteración	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	x_5	Lado derecho
0	Z	(0)	1	−3	−5	0	0	0	0
	x_3	(1)	0	1	0	1	0	0	4
	x_4	(2)	0	0	2	0	1	0	12
	x_5	(3)	0	3	2	0	0	1	18
1	Z	(0)	1						
	x_3	(1)	0						
	x_2	(2)	0	0	1	0	$\frac{1}{2}$	0	6
	x_5	(3)	0						

TABLA 4.6 Primeras dos tablas símplex del problema de la Wyndor Glass Co.

Iteración	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	x_5	Lado derecho
0	Z	(0)	1	−3	−5	0	0	0	0
	x_3	(1)	0	1	0	1	0	0	4
	x_4	(2)	0	0	2	0	1	0	12
	x_5	(3)	0	3	2	0	0	1	18
1	Z	(0)	1	−3	0	0	$\frac{5}{2}$	0	30
	x_3	(1)	0	1	0	1	0	0	4
	x_2	(2)	0	0	1	0	$\frac{1}{2}$	0	6
	x_5	(3)	0	3	0	0	−1	1	6

1. Divida el renglón pivote entre el número pivote. Use este *nuevo* renglón pivote en los pasos 2 y 3.
2. En los renglones (incluso el renglón 0) que tienen un coeficiente *negativo* en la columna pivote, se *suma* a este renglón el *producto* del valor absoluto de este coeficiente por el nuevo renglón pivote.
3. En el caso de los renglones que tienen un coeficiente *positivo* en la columna pivote, se les *resta* el *producto* de este coeficiente por el nuevo renglón pivote.

En el ejemplo: Debido a que x_2 sustituye a x_4 como variable básica, se necesita reproducir el patrón de la primera tabla de coeficientes de la columna de x_4 (0, 0, 1, 0) en la columna de x_2 de la segunda tabla símplex. Para comenzar, se divide el renglón pivote (renglón 2) entre el número pivote (2), de donde se obtiene el nuevo renglón 2 que aparece en la tabla 4.5. Después, se suma al renglón 0 el nuevo renglón 2 multiplicado por 5. Luego se resta del renglón 3 el nuevo renglón 2 multiplicado por 2 (o de manera equivalente, se resta del renglón 3 el renglón 2 *anterior*). Estos cálculos llevan a la nueva tabla símplex que se muestra en la tabla 4.6 de la iteración 1. Así, la nueva solución BF es (0, 6, 4, 0, 6), con $Z = 30$. Después se regresa a la prueba de optimalidad para verificar si la nueva solución BF es óptima. Como el nuevo renglón 0 todavía tiene un coeficiente negativo (−3 para x_1), la solución no es óptima, y se necesita por lo menos una iteración más.

Iteración 2 del ejemplo y la solución óptima que resulta

La segunda iteración comienza de nuevo en la segunda tabla símplex de la tabla 4.6 para encontrar la siguiente solución BF. Si se siguen las instrucciones de los pasos 1 y 2 se encuentra que x_1 es la variable básica entrante y x_5 la variable básica que sale, como se muestra en la tabla 4.7.

TABLA 4.7 Pasos 1 y 2 de la iteración 2 del problema de la Wyndor Glass Co.

Iteración	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	x_5	Lado derecho	Cociente
1	Z	(0)	1	−3	0	0	$\frac{5}{2}$	0	30	
	x_3	(1)	0	1	0	1	0	0	4	$\frac{4}{1} = 4$
	x_2	(2)	0	0	1	0	$\frac{1}{2}$	0	6	
	x_5	(3)	0	3	0	0	−1	1	6	$\frac{6}{3} = 2 \leftarrow$ mínimo

TABLA 4.8 Tabla símplex completa para el problema de la Wyndor Glass Co.

Iteración	Variable básica	Ec.	Coeficiente de:						Lado derecho
			Z	x_1	x_2	x_3	x_4	x_5	
0	Z	(0)	1	-3	-5	0	0	0	0
	x_3	(1)	0	1	0	1	0	0	4
	x_4	(2)	0	0	2	0	1	0	12
	x_5	(3)	0	3	2	0	0	1	18
1	Z	(0)	1	-3	0	0	$\frac{5}{2}$	0	30
	x_3	(1)	0	1	0	1	0	0	4
	x_2	(2)	0	0	1	0	$\frac{1}{2}$	0	6
	x_5	(3)	0	3	0	0	-1	1	6
2	Z	(0)	1	0	0	0	$\frac{3}{2}$	1	36
	x_3	(1)	0	0	0	1	$\frac{1}{3}$	$-\frac{1}{3}$	2
	x_2	(2)	0	0	1	0	$\frac{1}{2}$	0	6
	x_1	(3)	0	1	0	0	$-\frac{1}{3}$	$\frac{1}{3}$	2

En el paso 3 se divide el renglón pivote (renglón 3) de la tabla 4.7 entre el número pivote (3). Después, se suma al renglón 0 el nuevo renglón 3 multiplicado por 3. Luego, se resta el nuevo renglón 3 del renglón 1.

En la tabla 4.8 se tiene ahora el conjunto de tablas símplex completo. La nueva solución BF es (2, 6, 2, 0, 0), con $Z = 36$. Al hacer la prueba de optimalidad se encuentra que la solución es *óptima* porque no hay coeficientes negativos en el renglón 0, de manera que el algoritmo termina. En consecuencia, la solución óptima del problema de la Wyndor Glass Co. (antes de introducir variables de holgura) es $x_1 = 2$, $x_2 = 6$.

Ahora compare la tabla 4.8 con el trabajo que se hizo en la sección 4.3 para verificar que, en realidad, estas dos formas del método símplex son *equivalentes*. Después observe que la forma algebraica es mejor para entender la lógica que fundamenta el método símplex, pero la forma tabular organiza el trabajo de manera más conveniente y compacta. En general, de ahora en adelante se usará la forma tabular.

En el OR Tutor puede encontrar un **ejemplo adicional** de aplicación del método símplex en forma tabular. Vea la demostración llamada *Simplex Method—Tabular Form*. En la sección Worked Examples del sitio en internet del libro también se incluye **otro ejemplo** de este tipo.

4.5 ROMPIMIENTO DE EMPATES EN EL MÉTODO SÍMPLEX

Es posible que haya observado que en las dos secciones anteriores no se dijo qué hacer cuando las reglas de selección del método símplex no son suficientes para tomar una decisión clara, ya sea porque hay empates (valores iguales) o por otras ambigüedades parecidas. A continuación se estudiarán estos detalles.

Empate de la variable básica entrante

El paso 1 de cada iteración elige la variable no básica que tiene el coeficiente *negativo* con el *mayor valor absoluto* en la ecuación (0) actual como la variable básica entrante. Ahora suponga que dos o más variables no básicas tienen el coeficiente negativo más grande (en valor absoluto), es decir, que hay un empate. Por ejemplo, esto ocurriría en la primera iteración del problema de la Wyndor

TABLA 4.9 Tabla símplex inicial para el problema de la Wyndsor Glass Co. sin las dos últimas restricciones funcionales

Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	Lado derecho	Cociente
Z	(0)	1	-3	-5	0	0	
x_3	(1)	0	1	$\boxed{0}$	1	4	Ninguno

Donde $x_1 = 0$ y x_2 aumenta, $x_3 = 4 - 1x_1 - 0x_2 = 4 > 0$.

Glass Co. si se cambiara la función objetivo a $Z = 3x_1 + 3x_2$, con lo que la ecuación (0) inicial sería $Z - 3x_1 - 3x_2 = 0$. ¿Cómo debe romperse este empate?

La respuesta es que se puede elegir entre estos dos competidores de manera *arbitraria*. Tarde o temprano se llegará a la solución óptima, sin importar cuál de las variables empatadas se haya escogido, y no existe un método incontrovertible para predecir cuál conduce a la solución óptima con mayor rapidez. En este ejemplo, si se escoge x_1 como variable entrante, el método símplex alcanza la solución óptima (2, 6) en tres iteraciones, mientras que si se elige x_2 llega en dos.

Empate de la variable básica que sale: degeneración

Ahora suponga que el empate ocurre entre dos o más variables básicas cuando se elige la variable que sale en el paso 2 de una iteración. ¿Importa cuál se escoge? En teoría, sí, y en una forma crítica debido a lo que puede ocurrir en la siguiente sucesión de eventos. Primero, todas las variables empatadas se hacen cero al mismo tiempo cuando aumenta el valor de la variable entrante. Por lo tanto, aquellas que *no* se eligieron como variable básica saliente también tendrán un valor de cero en la nueva solución BF. (Las variables básicas con valor de *cero* se llaman **degeneradas**, y el mismo nombre se da a la solución BF correspondiente.) Segundo, si una de estas variables básicas degeneradas sigue con valor de cero hasta que se selecciona como variable básica que sale en una iteración posterior, la variable básica entrante deberá también quedar con valor de cero (ya que no puede crecer sin que la variable básica que sale se vuelva negativa), por lo que el valor de Z no cambiará. Tercero, si Z permanece igual en lugar de mejorar en cada iteración, el método símplex puede caer en un ciclo que repite la misma secuencia de soluciones en forma periódica, en lugar de aumentar en algún momento para llegar a la solución óptima. En realidad, se han construido ejemplos artificiales que se quedan atrapados en un ciclo perpetuo de este tipo.[10]

Por fortuna, aunque en teoría es posible que haya ciclos perpetuos, muy rara vez han ocurrido en problemas reales. Si ocurriera un ciclo siempre se puede salir de él al cambiar la elección de la variable básica que sale. Aún más, se han construido reglas especiales[11] para romper empates que siempre evitan los ciclos. Sin embargo, con frecuencia estas reglas se ignoran en las aplicaciones reales, y no se repetirán aquí. Para propósitos prácticos se recomienda romper los empates de modo arbitrario y seguir el proceso sin preocuparse de las variables degeneradas que puedan resultar.

Cuando no hay variable básica saliente: *Z* no acotada

Existe otra posibilidad en el paso 2 de una iteración, de la que no se ha hablado: aquella en la que *ninguna* variable califica como variable básica saliente.[12] Esta situación puede ocurrir si la variable básica entrante puede crecer de manera *indefinida* sin que *ninguna* de las variables básicas actuales adquiera valores negativos. En la forma tabular, esto significa que *todos* los coeficientes de la columna pivote (se excluye el renglón 0) son negativos o cero.

---

[10] Para obtener información adicional acerca de circular en un ciclo perpetuo, consulte la referencia J. A. J. Hall y K. I. M. McKinnon, "The Simplest Examples Where the Simplex Method Cycles and Conditions Where EXPAND Fails to Prevent Cycling", en *Mathematical Programming*, Series B, **100**(1): 135-150, mayo de 2004.

[11] Vea R. Bland, "New Finite Pivoting Rules for the Simplex Method", en *Mathematics of Operations Research*, **2**: 103-107, 1977.

[12] Note que el caso análogo (sin variable básica *entrante*) no puede ocurrir en el paso 1 de una iteración porque la prueba de optimalidad detendría el algoritmo antes, lo que indica que se ha alcanzado una solución óptima.

Como se ilustra en la tabla 4.9, esta situación surge en el ejemplo que se presentó en la figura 3.6. En él se pasaron por alto las dos últimas restricciones funcionales del problema de la Wyndor Glass Co., por lo cual no se incluyen en el modelo. Observe en la figura 3.6 que el valor de x_2 puede aumentar de manera indefinida (lo que hace que Z también lo haga) sin salir de la región factible. Después vea en la tabla 4.9 que x_2 es la variable básica entrante pero el único coeficiente en la columna pivote es cero. Como la prueba del cociente mínimo sólo utiliza coeficientes mayores que cero, no se cuenta con un cociente que proporcione una variable básica saliente.

La interpretación de una tabla símplex como la que se muestra en la tabla 4.9 sostiene que las restricciones no impiden el crecimiento indefinido de la función objetivo Z, de manera que el método símplex se detiene con el mensaje de que Z es *no acotada*. Debido a que ni siquiera la programación lineal ha descubierto la manera de lograr ganancias infinitas, el mensaje real en problemas prácticos es que se ha cometido un error. Tal vez el modelo esté mal formulado, ya sea por haber omitido una restricción relevante o por haberla establecido de modo incorrecto. De manera alternativa, pudo haber ocurrido un error en los cálculos.

Soluciones óptimas múltiples

En la sección 3.2 (en la definición de **solución óptima**) se mencionó que un problema puede tener más de una solución óptima. Este hecho se ejemplificó en la figura 3.5 cuando se cambió la función objetivo del problema de la Wyndor Glass Co. a $Z = 3x_1 + 2x_2$, de lo que resultó que todos los puntos sobre el segmento de recta entre (2, 6) y (4, 3) eran óptimos. En consecuencia, todas las soluciones óptimas son un *promedio ponderado* de estas dos soluciones FEV óptimas

$$(x_1, x_2) = w_1(2, 6) + w_2(4, 3).$$

donde los pesos w_1 y w_2 son números que satisfacen las relaciones

$$w_1 + w_2 = 1 \quad \text{y} \quad w_1 \geq 0, \quad w_2 \geq 0.$$

Por ejemplo, $w_1 = \frac{1}{3}$ y $w_2 = \frac{2}{3}$

$$(x_1, x_2) = \frac{1}{3}(2, 6) + \frac{2}{3}(4, 3) = \left(\frac{2}{3} + \frac{8}{3}, \quad \frac{6}{3} + \frac{6}{3}\right) = \left(\frac{10}{3}, \quad 4\right)$$

como una solución óptima.

En general, cualquier promedio ponderado de dos o más soluciones (vectores) donde los pesos son no negativos y suman 1 se llama **combinación convexa** de estas soluciones. Entonces, toda solución óptima del ejemplo es una combinación convexa de (2, 6) y (4, 3).

Este ejemplo es representativo de problemas con soluciones óptimas múltiples.

> Como se indica al final de la sección 3.2, *cualquier* problema de programación lineal con soluciones óptimas múltiples (y una región factible acotada) tiene al menos dos soluciones FEV que son óptimas. *Toda* solución óptima es una combinación convexa de estas soluciones FEV óptimas. En consecuencia, en la forma aumentada, toda solución óptima es una combinación convexa de las soluciones BF óptimas.

(Los problemas 4.5-5 y 4.5-6 son una guía para el razonamiento que fundamenta esta conclusión.)

El método símplex se detiene en forma automática cuando encuentra *una* solución BF óptima. Sin embargo, en muchas aplicaciones de programación lineal existen factores intangibles que no se incorporan al modelo y que pueden ser útiles para tomar decisiones significativas entre las soluciones óptimas alternativas. En esos casos, deben identificarse las otras soluciones óptimas. Como se indicó, esto requiere encontrar todas las otras soluciones BF óptimas, y entonces toda solución óptima es una combinación convexa de las soluciones BF óptimas.

Una vez que el método símplex encuentra una solución BF óptima se puede detectar si existen otras y, si así es, se encuentran como sigue:

> Siempre que un problema tiene más de una solución BF óptima, al menos una variable no básica tiene coeficiente cero en el renglón (0) final, de manera que si aumenta su valor, el valor de la función Z no cambia. Por lo tanto, estas otras soluciones BF óptimas se pueden identificar (si

■ **TABLA 4.10 Tabla símplex completa para obtener todas las soluciones BF óptimas con $c_2 = 2$, en el problemas de la Wyndor Glass Co.**

Iteración	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	x_5	Lado derecho	¿Solución óptima?
0	Z	(0)	1	−3	−2	0	0	0	0	No
	x_3	(1)	0	1	0	1	0	0	4	
	x_4	(2)	0	0	2	0	1	0	12	
	x_5	(3)	0	3	2	0	0	1	18	
1	Z	(0)	1	0	−2	3	0	0	12	No
	x_1	(1)	0	1	0	1	0	0	4	
	x_4	(2)	0	0	2	0	1	0	12	
	x_5	(3)	0	0	2	−3	0	1	6	
2	Z	(0)	1	0	0	**0**	0	1	18	Sí
	x_1	(1)	0	1	0	1	0	0	4	
	x_4	(2)	0	0	0	3	1	−1	6	
	x_2	(3)	0	0	1	$-\frac{3}{2}$	0	$\frac{1}{2}$	3	
Extra	Z	(0)	1	0	0	0	**0**	1	18	Sí
	x_1	(1)	0	1	0	0	$-\frac{1}{3}$	$\frac{1}{3}$	2	
	x_3	(2)	0	0	0	1	$\frac{1}{3}$	$-\frac{1}{3}$	2	
	x_2	(3)	0	0	1	0	$\frac{1}{2}$	0	6	

se desea) mediante iteraciones adicionales del método símplex, en las que cada vez se elige una variable no básica con coeficiente cero como variable básica entrante.[13]

A manera de ilustración, considere el caso anterior del problema de la Wyndor Glass Co., donde la función objetivo se cambia a $Z = 3x_1 + 2x_2$. En la tabla 4.10 se muestran las primeras tres tablas que obtiene el método símplex antes de detenerse con una solución básica factible óptima. No obstante, como una variable no básica (x_3) de esa iteración tiene coeficiente cero en el renglón 0, se realiza una iteración más en esa misma tabla para identificar la otra solución BF óptima. En consecuencia, las dos soluciones básicas factibles óptimas son (4, 3, 0, 6, 0) y (2, 6, 2, 0, 0), y ambas producen un valor de $Z = 18$. Observe que la última tabla símplex también tiene una variable *no básica* (x_4) con coeficiente cero en el renglón (0). Esta situación es inevitable porque las iteraciones adicionales no modifican el renglón 0, y cada una de las variables básicas que salen conserva su coeficiente cero. Si ahora se eligiera x_4 como variable básica entrante, sólo se regresaría a la tercera tabla símplex. (Verifique esto.) Por lo tanto, estas dos son las únicas soluciones BF óptimas, mientras que todas *las demás* son una combinación convexa de ellas.

$$(x_1, x_2, x_3, x_4, x_5) = w_1(2, 6, 2, 0, 0) + w_2(4, 3, 0, 6, 0),$$
$$w_1 + w_2 = 1, \qquad w_1 \geq 0, \qquad w_2 \geq 0.$$

4.6 ADAPTACIÓN A OTRAS FORMAS DE MODELO

Hasta ahora se han presentado los detalles del método símplex bajo el supuesto de que el problema se encuentra en nuestra forma estándar (maximizar Z sujeta a restricciones funcionales de la forma

[13] Si una de estas iteraciones no tiene una variable básica *saliente*, esto indica que la región factible es no acotada y la variable básica entrante puede crecer de manera indefinida sin cambiar el valor de Z.

≤ y restricciones de no negatividad sobre todas las variables) con $b_i \geq 0$ para toda $i = 1, 2, \ldots, m$. En esta sección se establecerá cómo hacer los ajustes requeridos a otras formas legítimas de modelos de programación lineal. Se verá que estos ajustes se pueden hacer en el paso inicial, y que el resto del método símplex se aplica exactamente como se aprendió.

El único problema serio que introducen las otras formas de restricciones funcionales (formas con = o ≥, o con lados derechos negativos) es identificar una *solución inicial básica factible*. Antes era muy sencillo encontrar esta solución inicial al hacer que las variables de holgura fueran las variables básicas iniciales, donde cada una era igual a la constante *no negativa* del lado derecho de la ecuación correspondiente. Ahora debe hacerse algo más. El enfoque estándar que se utiliza en estos casos es la **técnica de variables artificiales**. Ésta construye un *problema artificial* más conveniente mediante la introducción de una variable ficticia (llamada *variable artificial*) en cada restricción que lo requiera. Esta nueva variable se introduce sólo con el fin de que sea la variable básica inicial de esa ecuación. Las restricciones usuales de no negatividad también se aplican sobre estas variables y la función objetivo se modifica para que imponga una penalización exorbitante en el caso de que adquieran valores mayores que cero. De manera automática, las iteraciones del método símplex fuerzan a las variables artificiales a desaparecer (a volverse cero) una a una, hasta que todas quedan fuera de la solución; después de este paso se resuelve el problema *real*.

Para ilustrar la técnica de las variables artificiales, primero se considera el caso en que la única forma no estándar del problema es la presencia de una o más restricciones en forma de igualdad.

Restricciones en forma de igualdad

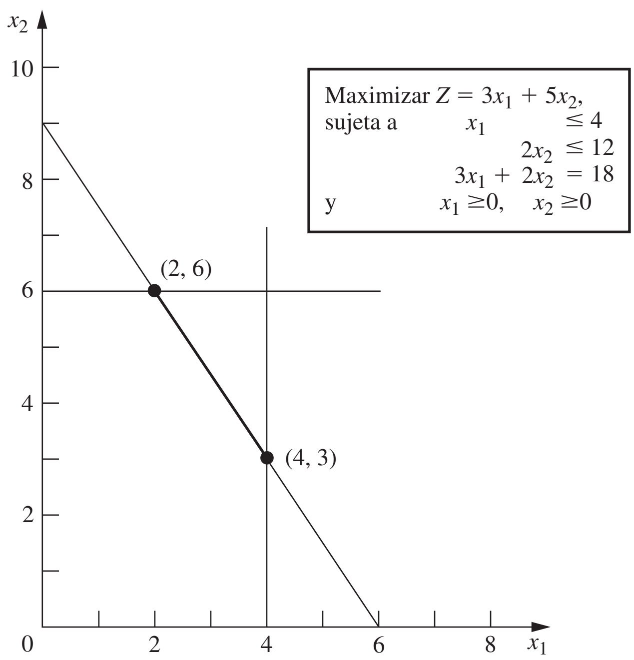
En realidad, cualquier restricción en forma de igualdad

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i$$

es equivalente a dos restricciones de desigualdad:

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leq b_i$$
$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \geq b_i.$$

■ **FIGURA 4.3**
Cuando la tercera restricción funcional se convierte en igualdad, la región factible del problema de la Wyndor Glass Co. se convierte en el segmento de recta entre (2, 6) y (4, 3).



Sin embargo, en lugar de hacer esta sustitución e incrementar con ello el número de restricciones, es más conveniente utilizar la técnica de la variable artificial que se ilustrará con el siguiente ejemplo.

Ejemplo. Suponga que se modifica el problema de la Wyndor Glass Co. de la sección 3.1, de manera que *se requiere* que la planta 3 se use a toda su capacidad. El único cambio que sufre el modelo de programación lineal es que la tercera restricción, $3x_1 + 2x_2 \leq 18$, se convierte en una restricción de igualdad

$$3x_1 + 2x_2 = 18,$$

con lo que el modelo completo se convierte en el que se muestra en la esquina superior derecha de la figura 4.3. Esta figura también muestra la región factible con tinta más oscura, y ahora consiste *nada más* en el segmento que conecta los puntos (2, 6) y (4, 3).

Después de introducir al sistema de ecuaciones las variables de holgura que todavía se necesitan para las restricciones de desigualdad, la forma aumentada del problema se convierte en

$$\begin{array}{lrcl}
(0) & Z - 3x_1 - 5x_2 & = & 0 \\
(1) & x_1 \qquad + x_3 \qquad & = & 4 \\
(2) & 2x_2 \qquad + x_4 & = & 12 \\
(3) & 3x_1 + 2x_2 \qquad\qquad & = & 18.
\end{array}$$

Desafortunadamente, estas ecuaciones no tienen una solución BF inicial obvia porque en la ecuación (3) ya no se tiene una variable de holgura para usar como variable básica inicial. Es necesario encontrar una solución BF inicial para comenzar con el método símplex.

Este obstáculo se vence de la siguiente manera.

Obtención de una solución BF inicial. El procedimiento es construir un **problema artificial** que tenga la misma solución óptima que el problema real, pero a este último se le deben hacer dos modificaciones.

1. Se aplica la **técnica de la variable artificial** mediante la introducción de una **variable artificial** *no negativa* (denotada por $\bar{x}_5$)[14] en la ecuación (3), como si fuera una variable de holgura

$$(3) \qquad 3x_1 + 2x_2 + \bar{x}_5 = 18.$$

2. Se asigna una *penalización enorme* al hecho de tener $\bar{x}_5 > 0$ para cambiar la función objetivo

$$Z = 3x_1 + 5x_2 \text{ a}$$
$$Z = 3x_1 + 5x_2 - M\bar{x}_5,$$

donde M representa en forma simbólica un número positivo *muy grande*. (Este método que fuerza a $\bar{x}_5$ hasta llegar a $\bar{x}_5 = 0$ en la solución óptima se llama **método de la gran *M*.**)

Ahora se encuentra la solución óptima del problema real con la aplicación del método símplex al problema artificial; la solución BF inicial es la siguiente:

Solución BF inicial
Variables no básicas: $x_1 = 0, \quad x_2 = 0$
Variables básicas: $x_3 = 4, \quad x_4 = 12, \quad \bar{x}_5 = 18.$

Como $\bar{x}_5$ asume el papel de la variable de holgura en la tercera restricción del problema artificial, esta restricción es equivalente a $3x_1 + 2x_2 \leq 18$ (igual que en el problema original de la Wyndor Glass Co. de la sección 3.1). A continuación se muestra el problema artificial que resulta (antes de aumentarlo) junto con el problema real.

[14] Se distinguirán siempre las variables artificiales mediante una barra sobre ellas.

El problema real

Maximizar $Z = 3x_1 + 5x_2$,

sujeta a

$$x_1 \leq 4$$
$$2x_2 \leq 12$$
$$3x_1 + 2x_2 = 18$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

El problema artificial

Definir $\bar{x}_5 = 18 - 3x_1 - 2x_2$.

Maximizar $Z = 3x_1 + 5x_2 - M\bar{x}_5$,

sujeta a

$$x_1 \leq 4$$
$$2x_2 \leq 12$$
$$3x_1 + 2x_2 \leq 18$$
$$(\text{así} \quad 3x_1 + 2x_2 + \bar{x}_5 = 18)$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad \bar{x}_5 \geq 0.$$

En consecuencia, igual que en la sección 3.1, la región factible de (x_1, x_2) en el problema artificial es la que se muestra en la figura 4.4. La única porción de esta región factible que coincide con la del problema original es cuando $\bar{x}_5 = 0$ (de manera que $3x_1 + 2x_2 = 18$).

La figura 4.4 muestra, además, el orden en el que el método símplex examina las soluciones FEV (o las soluciones BF después de aumentar el problema), donde los números en los círculos identifican qué iteración obtuvo esa solución. Observe que en este caso, el método símplex se mueve en sentido positivo (contrario al de las manecillas del reloj) mientras que en el problema original se movía en sentido negativo (vea la figura 4.2). La razón de esta diferencia es el término adicional $-M\bar{x}_5$ en la función objetivo del problema artificial.

Antes de aplicar el método símplex y comprobar que sigue el camino que se muestra en la figura 4.4, es necesario el siguiente paso de preparación.

Conversión de la ecuación (0) a la forma apropiada. El sistema de ecuaciones después de aumentar el problema artificial es

$$\begin{array}{llcl}
(0) & Z - 3x_1 - 5x_2 \qquad + M\bar{x}_5 & = & 0 \\
(1) & x_1 \qquad + \boldsymbol{x_3} & = & 4 \\
(2) & 2x_2 \qquad + \boldsymbol{x_4} & = & 12 \\
(3) & 3x_1 + 2x_2 \qquad + \quad \boldsymbol{\bar{x}_5} & = & 18
\end{array}$$

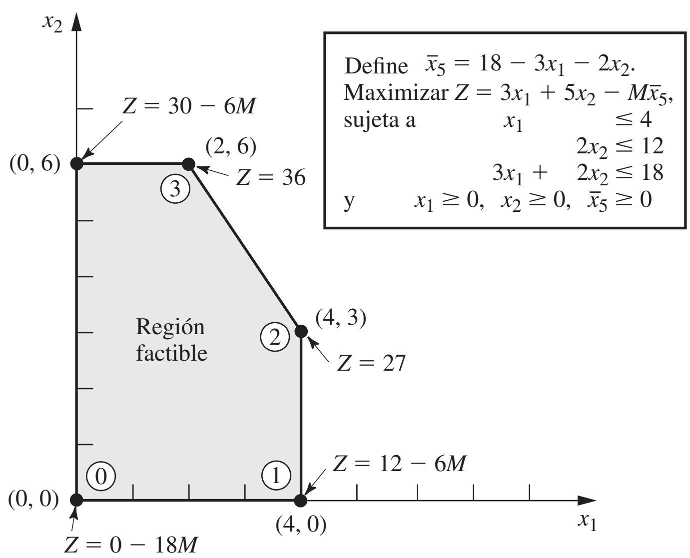


FIGURA 4.4
Esta gráfica muestra la región factible y la secuencia de soluciones FEV (⓪, ①, ②, ③) que examina el método símplex para el problema artificial que corresponde al problema real de la figura 4.3.

donde las variables básicas iniciales (x_3, x_4, $\bar{x}_5$) se muestran en **negritas**. Este sistema todavía no está en la forma apropiada de eliminación de Gauss porque el coeficiente de $\bar{x}_5$ es diferente de cero en la ecuación (0). Recuerde que todas las variables básicas deben eliminarse de la ecuación (0) con operaciones algebraicas antes de que el método símplex aplique la prueba de optimalidad o encuentre la variable básica entrante. Esta eliminación es necesaria para que el negativo del coeficiente de cada variable no básica proporcione la tasa a la que Z aumenta si esa variable no básica se incrementa a un valor mayor que 0 y se ajustan los valores de las variables básicas.

Para eliminar algebraicamente $\bar{x}_5$ de la ecuación (0) se resta de esta ecuación (0) la ecuación (3) multiplicada por M.

$$\begin{array}{rr} & Z - 3x_1 - 5x_2 + M\bar{x}_5 = \ \ 0 \\ & -M(3x_1 + 2x_2 + \ \ \bar{x}_5 = 18) \\ \hline \text{Nueva (0)} & Z - (3M + 3)x_1 - (2M + 5)x_2 = -18M. \end{array}$$

Aplicación del método símplex. Esta nueva ecuación (0) presenta a Z *sólo* en términos de las variables no básicas (x_1, x_2),

$$Z = -18M + (3M + 3)x_1 + (2M + 5)x_2.$$

Como $3M + 3 > 2M + 5$ (recuerde que M representa un número muy grande), si se aumenta x_1 Z crece a una tasa más rápida que al aumentar x_2, de manera que se elige x_1 como la variable básica entrante. Esto causa un movimiento de (0, 0) a (4, 0) en la iteración 1, como se muestra en la figura 4.4, y Z se incrementa en $4(3M + 3)$.

Las cantidades que incluyen el valor M nunca aparecen en el sistema de ecuaciones en otro renglón que no sea el (0), por lo que sólo tienen que tomarse en cuenta en la prueba de optimalidad y en el momento de determinar la variable básica entrante. Una manera de manejar estas cantidades es asignar a M algún valor numérico (muy grande) y trabajar con las cantidades que resulten en la ecuación (0) en la forma acostumbrada. Sin embargo, este enfoque puede acarrear errores de

TABLA 4.11 Conjunto completo de tablas símplex del problema de la figura 4.4

Iteración	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	$\bar{x}_5$	Lado derecho
0	Z	(0)	1	$-3M - 3$	$-2M - 5$	0	0	0	$-18M$
	x_3	(1)	0	1	0	1	0	0	4
	x_4	(2)	0	0	2	0	1	0	12
	$\bar{x}_5$	(3)	0	3	2	0	0	1	18
1	Z	(0)	1	0	$-2M - 5$	$3M + 3$	0	0	$-6M + 12$
	x_1	(1)	0	1	0	1	0	0	4
	x_4	(2)	0	0	2	0	1	0	12
	$\bar{x}_5$	(3)	0	0	2	-3	0	1	6
2	Z	(0)	1	0	0	$-\frac{9}{2}$	0	$M + \frac{5}{2}$	27
	x_1	(1)	0	1	0	1	0	0	4
	x_4	(2)	0	0	0	3	1	-1	6
	x_2	(3)	0	0	1	$-\frac{3}{2}$	0	$\frac{1}{2}$	3
3	Z	(0)	1	0	0	0	$\frac{3}{2}$	$M + 1$	36
	x_1	(1)	0	1	0	0	$-\frac{1}{3}$	$\frac{1}{3}$	2
	x_3	(2)	0	0	0	1	$\frac{1}{3}$	$-\frac{1}{3}$	2
	x_2	(3)	0	0	1	0	$\frac{1}{2}$	0	6

redondeo significativos que a su vez pueden invalidar la prueba de optimalidad. Por lo tanto, es mejor hacer lo que se acaba de explicar, esto es, expresar cada coeficiente de la ecuación (0) como una función lineal $aM + b$ de la cantidad *simbólica* M cuando se registra y actualiza por separado el valor numérico de: 1) el factor *multiplicativo* a y 2) el término *aditivo* b. Como se supone que M es tan grande que b es despreciable comparado con M cuando a es diferente de 0, las decisiones en la prueba de optimalidad y la elección de la variable básica entrante se hacen sólo con los valores de los factores *multiplicativos* en la forma usual. La única excepción ocurre cuando hay un empate que se rompe cuando se utilizan los factores *aditivos*.

En la tabla 4.11 se muestra la tabla símplex que resulta al aplicar esta técnica al ejemplo. Observe que la variable artificial $\bar{x}_5$ es una *variable básica* ($\bar{x}_5 > 0$) en las dos primeras tablas símplex y *no básica* ($\bar{x}_5 = 0$) en las últimas dos. En consecuencia, las dos primeras soluciones BF de este problema artificial son *no factibles* para el problema real mientras que las dos últimas son soluciones BF para el problema real.

Este ejemplo incluyó sólo una restricción de igualdad. Si un modelo de programación lineal tiene más, cada una debe manejarse de la misma manera. (Si el lado derecho es negativo, primero se multiplican ambos lados por -1.)

Lados derechos negativos

La técnica que acaba de presentarse para manejar una restricción de igualdad con lado derecho negativo (esto es, multiplicar ambos lados por -1) también se puede usar con las restricciones de desigualdad con lado derecho negativo. Si se multiplican ambos lados de una desigualdad por -1 se invierte el sentido de la desigualdad; es decir, $\leq$ cambia a $\geq$ o viceversa. Por ejemplo, al hacerlo con la restricción

$$x_1 - x_2 \leq -1 \qquad (\text{esto es, } x_1 \leq x_2 - 1)$$

se obtiene la restricción equivalente

$$-x_1 + x_2 \geq 1 \qquad (\text{esto es, } x_2 - 1 \geq x_1)$$

pero ahora el lado derecho es positivo. Cuando se tienen lados derechos no negativos para todas las restricciones funcionales, el método símplex puede comenzar porque (después de aumentar) estos lados derechos se convierten en los valores correspondientes a las *variables básicas iniciales*, que deben satisfacer las restricciones de no negatividad.

A continuación se estudiará cómo aumentar las restricciones del tipo $\geq$ como $-x_1 + x_2 \geq 1$, con la ayuda de la técnica de variables artificiales.

Restricciones funcionales de la forma $\geq$

Para ilustrar la manera en que la técnica de variables artificiales maneja las restricciones de la forma $\geq$ se usará el modelo del diseño de terapia de radiación para Mary, que se presentó en la sección 3.4. Por conveniencia se repite el modelo y se señala con un recuadro la restricción de interés en este caso.

Ejemplo de terapia de radiación

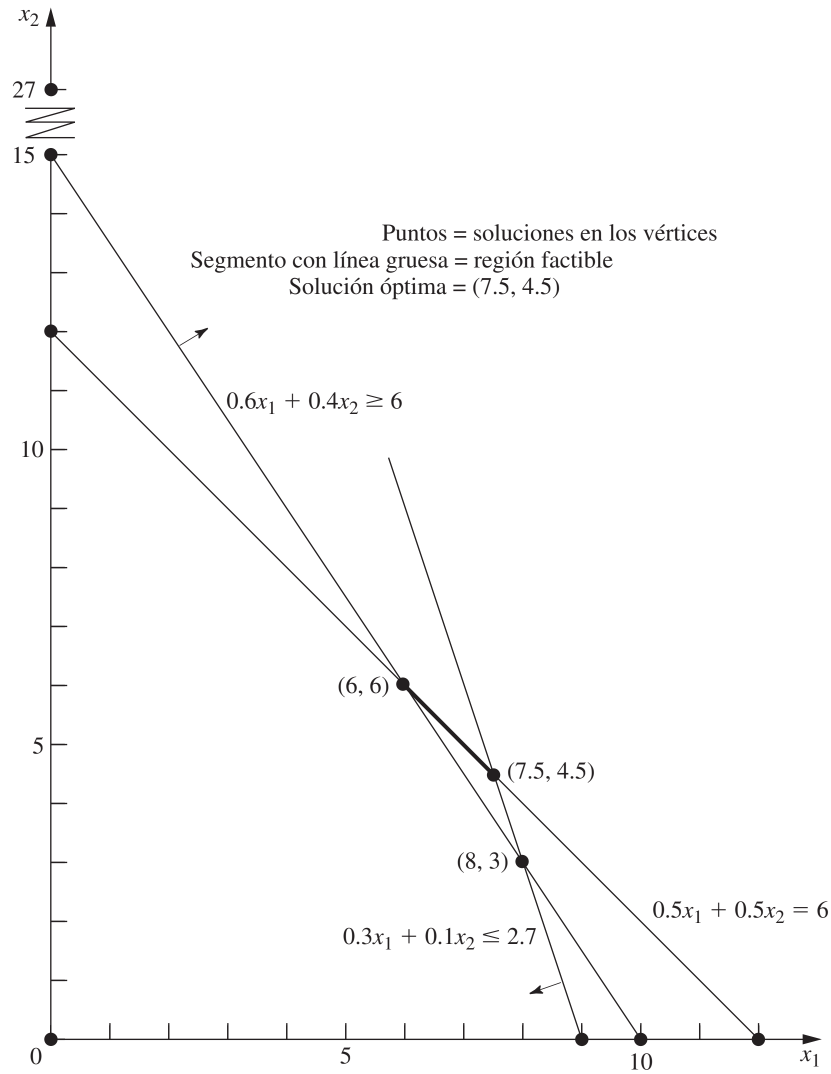
Minimizar $\quad Z = 0.4x_1 + 0.5x_2,$

sujeta a

$$0.3x_1 + 0.1x_2 \leq 2.7$$
$$0.5x_1 + 0.5x_2 = 6$$
$$\boxed{0.6x_1 + 0.4x_2 \geq 6}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$



■ **FIGURA 4.5**
Gráfica del ejemplo de terapia de radiación y sus soluciones en los vértices.

En la figura 4.5 se repite la solución gráfica de este ejemplo (que se muestra en la figura 3.12 con algunas diferencias). Las tres rectas de la figura, junto con los dos ejes constituyen las cinco fronteras de restricción del problema. Los puntos dibujados en la intersección de cada par de restricciones son las *soluciones en los vértices*. Las únicas dos *soluciones factibles* en un vértice son (6, 6) y (7.5, 4.5) y la región factible es el segmento que une estos dos puntos. La solución óptima es $(x_1, x_2) = (7.5, 4.5)$, con $Z = 5.25$.

Muy pronto se mostrará la manera en que el método símplex resuelve este problema a partir de la solución directa del problema artificial correspondiente. No obstante, primero se describirá cómo manejar la tercera restricción.

El enfoque involucra la introducción de *dos* variables: una variable de exceso x_5 (definida como $x_5 = 0.6x_1 + 0.4x_2 - 6$) y una variable artificial $\bar{x}_6$, como se verá en seguida.

$$
\begin{aligned}
& 0.6x_1 + 0.4x_2 && \geq 6 & \\
\rightarrow\quad & 0.6x_1 + 0.4x_2 - x_5 && = 6 & (x_5 \geq 0) \\
\rightarrow\quad & 0.6x_1 + 0.4x_2 - x_5 + \bar{x}_6 && = 6 & (x_5 \geq 0, \bar{x}_6 \geq 0).
\end{aligned}
$$

Aquí, x_5 se llama **variable de exceso** porque resta el excedente del lado izquierdo sobre el derecho para convertir la restricción de desigualdad en una de igualdad equivalente. Una vez que se logra esta conversión se introduce una variable artificial igual que para las restricciones de igualdad.

Una vez que se introduce la variable de holgura x_3 en la primera restricción, se inserta una variable artificial $\bar{x}_4$ en la segunda restricción y se aplica el método de la gran M; el problema artificial completo (en la forma aumentada) es

$$\begin{array}{llll} \text{Minimizar} & Z = 0.4x_1 + 0.5x_2 + M\bar{x}_4 + M\bar{x}_6, & & \\ \text{sujeta a} & 0.3x_1 + 0.1x_2 + x_3 & & = 2.7 \\ & 0.5x_1 + 0.5x_2 \qquad + \bar{x}_4 & & = 6 \\ & 0.6x_1 + 0.4x_2 \qquad\qquad - x_5 + \bar{x}_6 & & = 6 \end{array}$$

$$\text{y} \quad x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad \bar{x}_4 \geq 0, \quad x_5 \geq 0, \quad \bar{x}_6 \geq 0.$$

Observe que los coeficientes de las variables artificiales en la función objetivo son $+M$, en lugar de $-M$, porque ahora se debe minimizar Z. En consecuencia, aun cuando es posible que $\bar{x}_4 > 0$ y/o $\bar{x}_6 > 0$ sea una solución factible para el problema artificial, la unidad de penalización tan grande de $+M$ evita que esto ocurra en una solución óptima.

Como siempre, la introducción de variables artificiales amplía la región factible. Compare las restricciones sobre (x_1, x_2) en el problema real con las restricciones correspondientes del problema artificial.

Restricciones sobre (x_1, x_2) *del problema real*	*Restricciones sobre* (x_1, x_2) *del problema artificial*
$0.3x_1 + 0.1x_2 \leq 2.7$	$0.3x_1 + 0.1x_2 \leq 2.7$
$0.5x_1 + 0.5x_2 = 6$	$0.5x_1 + 0.5x_2 \leq 6$ (= se cumple si $\bar{x}_4 = 0$)
$0.6x_1 + 0.4x_2 \geq 6$	No hay restricción (excepto cuando $\bar{x}_6 = 0$)
$x_1 \geq 0, \quad x_2 \geq 0$	$x_1 \geq 0, \quad x_2 \geq 0$

La introducción de la variable artificial $\bar{x}_4$ para que asuma el papel de la variable de holgura en la segunda restricción, permite valores de (x_1, x_2) *por debajo* de la recta $0.5x_1 + 0.5x_2 = 6$ de la figura 4.5. La introducción de x_5 y $\bar{x}_6$ en la tercera restricción del problema real (y el movimiento de estas variables al lado derecho) lleva a la ecuación

$$0.6x_1 + 0.4x_2 = 6 + x_5 - \bar{x}_6.$$

Debido a que la única restricción sobre x_5 y $\bar{x}_6$ es que sean no negativas, su diferencia $x_5 - \bar{x}_6$ puede ser un número positivo o negativo. Entonces, $0.6x_1 + 0.4x_2$ puede tomar cualquier valor, lo que tiene el efecto de eliminar la tercera restricción del problema artificial y permitir puntos en los dos lados de la recta $0.6x_1 + 0.4x_2 = 6$ en la figura 4.5. (Se conserva la tercera restricción del sistema de ecuaciones porque más adelante volverá a ser relevante, después de que el método de la gran M lleve a $\bar{x}_6$ a cero.) En consecuencia, la región factible del problema artificial es el poliedro completo de la figura 4.5 cuyos vértices son (0, 0), (9, 0), (7.5, 4.5) y (0, 12).

Como ahora el origen es factible para el problema artificial, el método símplex comienza con (0, 0) como la solución FEV inicial, es decir, con $(x_1, x_2, x_3, \bar{x}_4, x_5, \bar{x}_6) = (0, 0, 2.7, 6, 0, 6)$ como solución BF inicial. (El propósito de crear el problema artificial es hacer que el origen sea factible para tener un punto de partida conveniente para el método símplex.) Después se seguirá la trayectoria del método símplex desde el origen hasta la solución óptima para los dos problemas, el artificial y el real. Pero primero, ¿cómo maneja el método símplex la *minimización*?

Minimización

Una manera directa de minimizar Z con el método símplex es cambiar los roles de los coeficientes negativos y positivos en el renglón 0, tanto para la prueba de optimalidad como para el paso 1 de una iteración. Sin embargo, en lugar de cambiar las instrucciones del método símplex para este caso, se presentará una manera sencilla de convertir cualquier problema de minimización en un problema equivalente de maximización:

$$\textit{Minimizar} \qquad Z = \sum_{j=1}^{n} c_j x_j$$

es equivalente a

$$\textit{maximizar} \qquad -Z = \sum_{j=1}^{n} (-c_j) x_j;$$

es decir, las dos formulaciones llevan a la(s) misma(s) solución(es) óptima(s).

Las dos formulaciones son equivalentes porque mientras más pequeña es Z, más grande es $-Z$; entonces, la solución que da el *menor* valor para Z dentro de la región factible, también debe dar el *mayor* valor para $-Z$ en esta región.

Por lo tanto, en el ejemplo de terapia de radiación se debe hacer el siguiente cambio en la formulación:

$$\begin{array}{lll} & \text{Minimizar} & Z = \ \ 0.4x_1 + 0.5x_2 \\ \rightarrow & \text{Maximizar} & -Z = -0.4x_1 - 0.5x_2. \end{array}$$

Después de introducir las variables artificiales $\bar{x}_4$ y $\bar{x}_6$ y de aplicar el método de la gran M, la conversión correspondiente es

$$\begin{array}{lll} & \text{Minimizar} & Z = \ \ 0.4x_1 + 0.5x_2 + M\bar{x}_4 + M\bar{x}_6 \\ \rightarrow & \text{Maximizar} & -Z = -0.4x_1 - 0.5x_2 - M\bar{x}_4 - M\bar{x}_6. \end{array}$$

Solución del ejemplo de terapia de radiación

El ejemplo está casi listo para que se le aplique el método símplex. Si se usa la forma de maximización que se acaba de obtener, el sistema de ecuaciones completo es

$$\begin{array}{lrcl} (0) & -Z + 0.4x_1 + 0.5x_2 & + M\bar{\boldsymbol{x}}_4 \qquad + M\bar{\boldsymbol{x}}_6 & = 0 \\ (1) & 0.3x_1 + 0.1x_2 + \boldsymbol{x}_3 & & = 2.7 \\ (2) & 0.5x_1 + 0.5x_2 & + \ \ \bar{\boldsymbol{x}}_4 & = 6 \\ (3) & 0.6x_1 + 0.4x_2 & - x_5 + \ \ \bar{\boldsymbol{x}}_6 & = 6. \end{array}$$

Las variables básicas $(x_3, \bar{x}_4, \bar{x}_6)$ de la solución BF inicial (para este problema artificial) se muestran en **negritas**.

Observe que este sistema de ecuaciones todavía no está en la forma apropiada de eliminación gaussiana para iniciar el método símplex, puesto que todavía deben eliminarse las variables básicas $\bar{x}_4$ y $\bar{x}_6$ de la ecuación (0) de manera algebraica. Como $\bar{x}_4$ y $\bar{x}_6$ tienen coeficiente M se tienen que restar de la ecuación (0), las ecuaciones (2) *y* (3), *ambas* multiplicadas por M. A continuación se resumen los cálculos de todos los coeficientes (y los lados derechos), en donde los vectores son los renglones relevantes de la tabla símplex correspondiente al sistema de ecuaciones anterior.

Renglón 0:

$$\begin{array}{lrrrrrrr} & [0.4, & 0.5, & 0, & M, & 0, & M, & 0] \\ & -M[0.5, & 0.5, & 0, & 1, & 0, & 0, & 6] \\ & -M[0.6, & 0.4, & 0, & 0, & -1, & 1, & 6] \\ \hline \text{Nuevo renglón} & 0 = [-1.1M + 0.4, & -0.9M + 0.5, & 0, & 0, & M, & 0, & -12M] \end{array}$$

La tabla símplex inicial que resulta, lista para comenzar el método símplex, se muestra en la tabla 4.12. Al aplicar el método símplex en la forma acostumbrada se obtiene la secuencia de tablas símplex que aparecen en el resto de la tabla 4.12. En cuanto a la prueba de optimalidad y la elección de la variable básica entrante en cada iteración, las cantidades que incluyen M se tratan tal como se explicó para la tabla 4.11. En particular, siempre que M está presente, sólo se usa su factor multiplicativo a menos que haya un empate, en cuyo caso el empate se rompe con los términos aditivos correspondientes. Un empate de este tipo ocurre en la última selección de la variable básica entrante (vea la penúltima tabla símplex) en donde los coeficientes de x_3 y x_5 en el renglón 0 tienen el mismo factor multiplicativo de $-\frac{5}{3}$. Cuando se comparan los términos aditivos, $\frac{11}{6} < \frac{7}{3}$, se selecciona x_5 como la variable básica entrante.

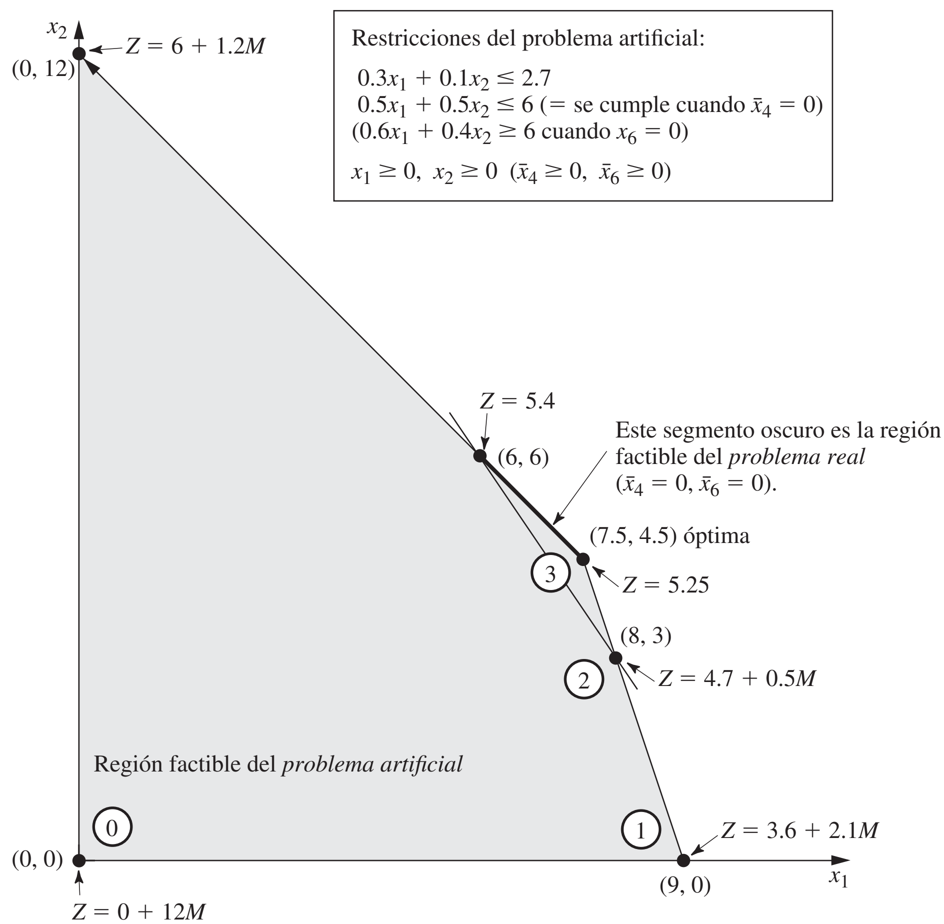
■ **TABLA 4.12** El método de la gran *M* en el ejemplo de terapia de radiación

Iteración	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	$\bar{x}_4$	x_5	$\bar{x}_6$	Lado derecho
0	Z	(0)	-1	$-1.1M + 0.4$	$-0.9M + 0.5$	0	0	M	0	$-12M$
	x_3	(1)	0	0.3	0.1	1	0	0	0	2.7
	$\bar{x}_4$	(2)	0	0.5	0.5	0	1	0	0	6
	$\bar{x}_6$	(3)	0	0.6	0.4	0	0	-1	1	6
1	Z	(0)	-1	0	$-\frac{16}{30}M + \frac{11}{30}$	$\frac{11}{3}M - \frac{4}{3}$	0	M	0	$-2.1M - 3.6$
	x_1	(1)	0	1	$\frac{1}{3}$	$\frac{10}{3}$	0	0	0	9
	$\bar{x}_4$	(2)	0	0	$\frac{1}{3}$	$-\frac{5}{3}$	1	0	0	1.5
	$\bar{x}_6$	(3)	0	0	0.2	-2	0	-1	1	0.6
2	Z	(0)	-1	0	0	$-\frac{5}{3}M + \frac{7}{3}$	0	$-\frac{5}{3}M + \frac{11}{6}$	$\frac{8}{3}M - \frac{11}{6}$	$-0.5M - 4.7$
	x_1	(1)	0	1	0	$\frac{20}{3}$	0	$\frac{5}{3}$	$-\frac{5}{3}$	8
	$\bar{x}_4$	(2)	0	0	0	$\frac{5}{3}$	1	$\frac{5}{3}$	$-\frac{5}{3}$	0.5
	x_2	(3)	0	0	1	-10	0	-5	5	3
3	Z	(0)	-1	0	0	0.5	$M - 1.1$	0	M	-5.25
	x_1	(1)	0	1	0	5	-1	0	0	7.5
	x_5	(2)	0	0	0	1	0.6	1	-1	0.3
	x_2	(3)	0	0	1	-5	3	0	0	4.5

Observe en la tabla 4.12 la progresión de valores de las variables artificiales $\bar{x}_4$ y $\bar{x}_6$ y de Z. Se comienza con valores grandes, $\bar{x}_4 = 6$ y $\bar{x}_6 = 6$, con $Z = 12M$ ($-Z = -12M$). La primera iteración reduce de manera considerable estos valores. El método de la M logra hacer que $\bar{x}_6$ sea cero (como una nueva variable no básica) en la segunda iteración y en la siguiente hace lo mismo con $\bar{x}_4$. Con $\bar{x}_4 = 0$ y $\bar{x}_6 = 0$ se garantiza que la solución básica que se obtuvo en la última tabla símplex es factible para el problema real. Debido a que cumple con la prueba de optimalidad, también es óptima.

Ahora se puede ver lo que el método de la gran M ha hecho en la gráfica de la figura 4.6. Al iniciar, la región factible del problema artificial tiene cuatro soluciones FEV, (0, 0), (9, 0), (0, 12) y (7.5, 4.5), y después sustituye las primeras tres con dos soluciones FEV nuevas, (8, 3), (6, 6), después de que $\bar{x}_6$ decrece hasta $\bar{x}_6 = 0$ de manera que $0.6x_1 + 0.4x_2 \geq 6$ se convierte en una restricción adicional. (Observe que las tres soluciones FEV sustituidas, (0, 0), (9, 0) y (0, 12), eran, en realidad, soluciones *no factibles* en los vértices del problema real que se presentó en la figura 4.5.) Si se comienza con el origen como una solución FEV inicial conveniente para el problema artificial, el proceso se mueve por la frontera a las otras tres soluciones FEV, (9, 0), (8, 3) y (7.5, 4.5). La última de éstas es la primera que también es factible para el problema real. Por coincidencia, esta primera solución factible resulta óptima, por lo que no se necesitan más iteraciones.

En el caso de otros problemas con variables artificiales, puede ser necesario realizar iteraciones adicionales para llegar a una solución óptima después de obtener la primera solución factible del problema real. (Éste fue el caso del ejemplo resuelto en la tabla 4.11.) De esta forma, puede pensarse que el método de la gran M tiene dos fases. En la *primera fase*, todas las variables artificiales se hacen cero (debido a la penalización de M por unidad al ser mayores que cero) con el fin de obtener una solución básica factible inicial para el problema *real*. En la *segunda fase* todas las variables artificiales se mantienen en cero (por esta misma penalización), mientras que el método símplex genera una secuencia de soluciones BF que llevan a la solución óptima. El *método de las dos fases* que se describe a continuación es un procedimiento directo para realizar estas dos fases sin siquiera introducir la M de una manera explícita.



■ **FIGURA 4.6**
La gráfica muestra la región factible y la secuencia de soluciones FEV (⓪, ①, ②, ③) que se examinaron por el método símplex (con el método de la gran *M*) del problema artificial correspondiente al problema real de la figura 4.5.

Método de las dos fases

En el ejemplo de terapia de radiación que se acaba de resolver en la tabla 4.12, recuerde que la función objetivo real es

Problema real: Minimizar $Z = 0.4x_1 + 0.5x_2$.

Sin embargo, el método de la gran *M* utiliza la siguiente función objetivo (o su equivalente en forma de maximización) en todo el procedimiento:

Método de la gran M: Minimizar $Z = 0.4x_1 + 0.5x_2 + M\bar{x}_4 + M\bar{x}_6$.

Como los dos primeros coeficientes son despreciables comparados con *M*, el método de las dos fases puede eliminar la *M* si se usan las siguientes dos funciones objetivo que definen a *Z* de manera completamente diferente.

Método de las dos fases:

Fase 1: Minimizar $Z = \bar{x}_4 + \bar{x}_6$ (hasta $\bar{x}_4 = 0$, $\bar{x}_6 = 0$).
Fase 2: Minimizar $Z = 0.4x_1 + 0.5x_2$ (con $\bar{x}_4 = 0$, $\bar{x}_6 = 0$).

La función objetivo de la fase 1 se obtiene si se divide la función objetivo del método de la gran *M* entre *M* y eliminan los términos despreciables. Como la fase 1 concluye cuando se obtiene una solución BF para el problema real (aquella en la que $\bar{x}_4 = 0$ y $\bar{x}_6 = 0$), esta solución se usa como la solución BF *inicial* para aplicar el método símplex al problema real (con su función objetivo) en la fase 2.

Antes de resolver el ejemplo por este método se hará un resumen de sus características generales.

Resumen del método de las dos fases. *Paso inicial*: se revisan las restricciones del problema original y se introducen variables artificiales según se necesite para obtener una solución BF inicial obvia para el *problema artificial*.

Fase 1: El objetivo de esta fase es encontrar una solución BF para el *problema real*. Para hacerlo, se debe,

Minimizar $Z = \sum$ de variables artificiales, sujeta a las restricciones revisadas.

La solución óptima que se obtiene para este problema (con $Z = 0$) será una solución BF para el problema real.

Fase 2: El objetivo de esta fase es encontrar una *solución óptima* para el problema real. Como las variables artificiales no son parte del problema real, ahora se pueden eliminar (de cualquier manera todas tienen valor de cero).[15] Se comienza con la solución BF que se obtuvo al final de la fase 1 y se usa el método símplex para resolver el problema real.

A continuación se resumen los problemas que deben resolverse por el método símplex en las fases respectivas del ejemplo.

Problema de la fase 1 (ejemplo de terapia de radiación):

Minimizar $\quad Z = \bar{x}_4 + \bar{x}_6,$

sujeta a

$$
\begin{aligned}
0.3x_1 + 0.1x_2 + x_3 \qquad\qquad\quad &= 2.7 \\
0.5x_1 + 0.5x_2 \qquad + \bar{x}_4 \qquad\quad &= 6 \\
0.6x_1 + 0.4x_2 \qquad\qquad - x_5 + \bar{x}_6 &= 6
\end{aligned}
$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad \bar{x}_4 \geq 0, \quad x_5 \geq 0, \quad \bar{x}_6 \geq 0.$$

Problema de la fase 2 (ejemplo de terapia de radiación):

Minimizar $\quad Z = 0.4x_1 + 0.5x_2,$

sujeta a

$$
\begin{aligned}
0.3x_1 + 0.1x_2 + x_3 \quad &= 2.7 \\
0.5x_1 + 0.5x_2 \qquad &= 6 \\
0.6x_1 + 0.4x_2 \quad - x_5 &= 6
\end{aligned}
$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_5 \geq 0.$$

Las únicas diferencias entre estos dos problemas se encuentran en la función objetivo y en la inclusión (fase 1) o exclusión (fase 2) de las variables artificiales $\bar{x}_4$ y $\bar{x}_6$. Sin las variables artificiales, el problema de la fase 2 no tiene una *solución BF inicial* obvia. El único propósito de resolver el problema de la fase 1 es obtener una solución BF con $\bar{x}_4 = 0$ y $\bar{x}_6 = 0$ que se pueda usar como la solución BF inicial para la fase 2.

La tabla 4.13 muestra el resultado de aplicar el método símplex a este problema en la fase 1. [El renglón 0 de la tabla símplex inicial se obtiene al convertir minimizar $Z = \bar{x}_4 + \bar{x}_6$ en maximizar $(-Z) = -\bar{x}_4 - \bar{x}_6$ y después usar *operaciones elementales con renglones* para eliminar las varia-

[15] Omitimos otras tres posibilidades: 1) variables artificiales > 0 (que se estudian en la subsección siguiente), 2) variables artificiales que son variables básicas degeneradas y 3) conservar las variables artificiales como variables no básicas en la fase 2 (sin permitir que se conviertan en básicas) como ayuda en el análisis de posoptimización subsecuente. El tutorial IOR le permitirá analizar dichas posibilidades.

■ **TABLA 4.13 Fase 1 del método de las dos fases en el ejemplo de la terapia de radiación**

Iteración	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	$\bar{x}_4$	x_5	$\bar{x}_6$	Lado derecho
0	Z	(0)	−1	−1.1	−0.9	0	0	1	0	−12
	x_3	(1)	0	0.3	0.1	1	0	0	0	2.7
	$\bar{x}_4$	(2)	0	0.5	0.5	0	1	0	0	6
	$\bar{x}_6$	(3)	0	0.6	0.4	0	0	−1	1	6
1	Z	(0)	−1	0	$-\frac{16}{30}$	$\frac{11}{3}$	0	1	0	−2.1
	x_1	(1)	0	1	$\frac{1}{3}$	$\frac{10}{3}$	0	0	0	9
	$\bar{x}_4$	(2)	0	0	$\frac{1}{3}$	$-\frac{5}{3}$	1	0	0	1.5
	$\bar{x}_6$	(3)	0	0	0.2	−2	0	−1	1	0.6
2	Z	(0)	−1	0	0	$-\frac{5}{3}$	0	$-\frac{5}{3}$	$\frac{8}{3}$	−0.5
	x_1	(1)	0	1	0	$\frac{20}{3}$	0	$\frac{5}{3}$	$-\frac{5}{3}$	8
	$\bar{x}_4$	(2)	0	0	0	$\frac{5}{3}$	1	$\frac{5}{3}$	$-\frac{5}{3}$	0.5
	x_2	(3)	0	0	1	−10	0	−5	5	3
3	Z	(0)	−1	0	0	0	1	0	1	0
	x_1	(1)	0	1	0	0	−4	−5	5	6
	x_3	(2)	0	0	0	1	$\frac{3}{5}$	1	−1	0.3
	x_2	(3)	0	0	1	0	6	5	−5	6

bles básicas $\bar{x}_4$ y $\bar{x}_6$ de $-Z + \bar{x}_4 + \bar{x}_6 = 0$.] En la penúltima tabla símplex existe un empate en la *variable básica entrante* entre x_3 y x_5, que se rompe de manera arbitraria a favor de x_3. La solución que se obtiene al final de la fase 1 es, entonces, $(x_1, x_2, x_3, \bar{x}_4, x_5, \bar{x}_6) = (6, 6, 0.3, 0, 0, 0)$ o después de eliminar $\bar{x}_4$ y $\bar{x}_6$, $(x_1, x_2, x_3, x_5) = (6, 6, 0.3, 0)$.

Según se afirmó en el resumen, esta solución de la fase 1 es, sin duda, una solución BF para el problema *real* (problema de la fase 2) puesto que es la solución (después de hacer $x_5 = 0$) del sistema de ecuaciones que consiste en las tres restricciones funcionales del problema de la fase 2. De hecho, después de eliminar las columnas de $\bar{x}_4$ y $\bar{x}_6$ al igual que el renglón 0 en cada iteración, la tabla 4.13 muestra una manera de utilizar la eliminación gaussiana para resolver este sistema de ecuaciones mediante su reducción a la forma que tiene en la tabla símplex final.

La tabla 4.14 muestra la preparación para iniciar la fase 2 después de completar la fase 1. Se comienza con la última tabla símplex de la tabla 4.13, se eliminan las variables artificiales ($\bar{x}_4$ y $\bar{x}_6$), se sustituye la función objetivo de la fase 2 ($-Z = -0.4x_1 - 0.5x_2$ en la forma de maximización) en el renglón 0 y después se restablece la forma apropiada de eliminación gaussiana (con la eliminación algebraica de las variables básicas x_1 y x_2 del renglón 0). De esta forma, el renglón 0 de la tabla símplex se obtiene mediante las siguientes *operaciones elementales con renglones* en la penúltima tabla símplex: se restan del renglón 0, el renglón 1 multiplicado por 0.4 y el renglón 3 multiplicado por 0.5. Observe que los renglones 1 y 3 no cambian excepto por la eliminación de las dos columnas. Los únicos ajustes ocurren en el renglón 0 a fin de sustituir la función objetivo de la fase 1 por la función objetivo de la fase 2.

La última tabla símplex de la tabla 4.14 es la tabla símplex inicial para aplicar el método símplex al problema de la fase 2, como se muestra al inicio de la tabla 4.15. Una sola iteración conduce a la solución óptima que se muestra en la segunda tabla símplex: $(x_1, x_2, x_3, x_5) = (7.5, 4.5, 0, 0.3)$. Ésta es la solución óptima deseada del problema real que interesa más que el problema artificial que se construyó en la fase 1.

TABLA 4.14 Preparación para la fase 2 en el ejemplo de terapia de radiación

	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	$\bar{x}_4$	x_5	$\bar{x}_6$	Lado derecho
Tabla símplex final fase 1	Z	(0)	−1	0	0	0	1	0	1	0
	x_1	(1)	0	1	0	0	−4	−5	5	6
	x_3	(2)	0	0	0	1	$\frac{3}{5}$	1	−1	0.3
	x_2	(3)	0	0	1	0	6	5	−5	6
Se eliminan $\bar{x}_4$ y $\bar{x}_6$	Z	(0)	−1	0	0	0		0		0
	x_1	(1)	0	1	0	0		−5		6
	x_3	(2)	0	0	0	1		1		0.3
	x_2	(3)	0	0	1	0		5		6
Se sustituye la función objetivo de la fase 2	Z	(0)	−1	0.4	0.5	0		0		0
	x_1	(1)	0	1	0	0		−5		6
	x_3	(2)	0	0	0	1		1		0.3
	x_2	(3)	0	0	1	0		5		6
Se restablece la forma apropiada de eliminación gaussiana	Z	(0)	−1	0	0	0		−0.5		−5.4
	x_1	(1)	0	1	0	0		−5		6
	x_3	(2)	0	0	0	1		1		0.3
	x_2	(3)	0	0	1	0		5		6

Ahora observe lo que el método de las dos fases ha hecho gráficamente en la figura 4.7. A partir del origen, la fase 1 examina un total de cuatro soluciones FEV para el problema artificial. En realidad, las primeras tres eran soluciones no factibles en los vértices para el problema real que se presentó en la figura 4.5. La cuarta solución FEV, en (6,6), es la primera que también es factible para el problema real, de manera que se convierte en la solución FEV inicial de la fase 2. Después de una iteración se obtiene la solución FEV óptima en (7.5, 4.5).

Si el empate en la variable básica entrante que surgió en la penúltima tabla símplex de la tabla 4.13 se hubiera roto de otra manera, la fase 1 habría ido directamente de (8, 3) a (7.5, 4.5). Después de utilizar (7.5, 4.5) para establecer la tabla símplex inicial de la fase 2, la *prueba de optimalidad* habría revelado que esta solución era óptima y no se habría realizado ninguna iteración.

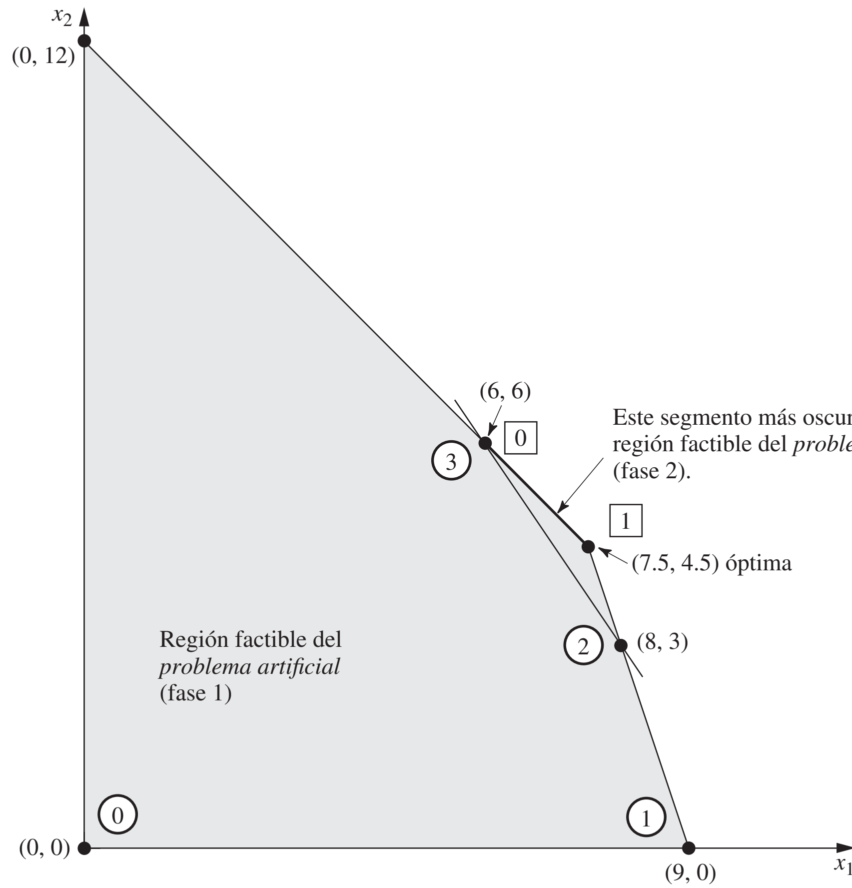
Resulta interesante comparar los métodos de la gran M y de las dos fases. Se comenzará por sus funciones objetivos.

Método de la gran M:

Minimizar $\quad Z = 0.4x_1 + 0.5x_2 + M\bar{x}_4 + M\bar{x}_6.$

TABLA 4.15 Fase 2 del método de las dos fases en el ejemplo de terapia de radiación

Iteración	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_5	Lado derecho
0	Z	(0)	−1	0	0	0	−0.5	−5.4
	x_1	(1)	0	1	0	0	−5	6
	x_3	(2)	0	0	0	1	1	0.3
	x_2	(3)	0	0	1	0	5	6
1	Z	(0)	−1	0	0	0.5	0	−5.25
	x_1	(1)	0	1	0	5	0	7.5
	x_5	(2)	0	0	0	1	1	0.3
	x_2	(3)	0	0	1	−5	0	4.5



■ **FIGURA 4.7**
La gráfica muestra la secuencia de soluciones FEV de la fase 1 (⓪, ①, ②, ③) y después de la fase 2 ([0], [1]) cuando se aplica el método de las dos fases al ejemplo de terapia de radiación.

Método de las dos fases:

Fase 1: Minimizar $Z = \bar{x}_4 + \bar{x}_6$.
Fase 2: Minimizar $Z = 0.4x_1 + 0.5x_2$.

Dado que los términos $M\bar{x}_4$ y $M\bar{x}_6$ dominan a los términos $0.4x_1$ y $0.5x_2$ en la función objetivo del método de la gran *M*, esta función objetivo es esencialmente equivalente a la de la fase 1 siempre que $\bar{x}_4$ y/o $\bar{x}_6$ sean mayores que cero. Entonces, cuando $\bar{x}_4 = 0$ y $\bar{x}_6 = 0$, la función objetivo del método de la gran *M* se vuelve completamente equivalente a la función objetivo de la fase 2.

Debido a estas equivalencias virtuales de las funciones objetivo, el método de la gran *M* y el de dos fases tienen casi siempre la misma secuencia de soluciones básicas factibles. La única excepción posible ocurre cuando existe un empate en la variable básica entrante en la fase 1 del método de las dos fases, como sucedió en la tercera tabla símplex de la tabla 4.13. Observe que las primeras tres tablas símplex de las tablas 4.12 y 4.13 son casi idénticas, pues la única diferencia es que los factores multiplicativos de *M* de la tabla 4.12 se convierten en cantidades únicas en los puntos correspondientes de la tabla 4.13. En consecuencia, no se contaba con los factores aditivos que rompieron el empate de la variable básica entrante en la tercera tabla símplex de la tabla 4.12 para romper este mismo empate en la tabla 4.13. El resultado en este ejemplo fue una iteración adicional en el método de las dos fases, aunque, en general, la ventaja de contar con los factores aditivos es mínima.

El método de las dos fases sigue los pasos del método de la gran *M*, pero en la fase 1 utiliza sólo los factores multiplicativos, mientras que en la fase 2 elimina las variables artificiales. (El método de la gran *M* puede combinar los factores multiplicativos y aditivos y asignar un valor muy grande a *M*, procedimiento que podría crear problemas con la inestabilidad numérica.) Por estas razones es común que cuando se trate de paquetes de computadora se utilice el método de las dos fases.

TABLA 4.16 **El método de la gran *M* para la revisión del ejemplo de terapia de radiación que no tiene soluciones factibles**

Iteración	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	$\bar{x}_4$	x_5	$\bar{x}_6$	Lado derecho
0	Z	(0)	−1	$-1.1M + 0.4$	$-0.9M + 0.5$	0	0	M	0	$-12M$
	x_3	(1)	0	0.3	0.1	1	0	0	0	1.8
	$\bar{x}_4$	(2)	0	0.5	0.5	0	1	0	0	6
	$\bar{x}_6$	(3)	0	0.6	0.4	0	0	−1	1	6
1	Z	(0)	−1	0	$-\frac{16}{30}M + \frac{11}{30}$	$\frac{11}{3}M - \frac{4}{3}$	0	M	0	$-5.4M - 2.4$
	x_1	(1)	0	1	$\frac{1}{3}$	$\frac{10}{3}$	0	0	0	6
	$\bar{x}_4$	(2)	0	0	$\frac{1}{3}$	$-\frac{5}{3}$	1	0	0	3
	$\bar{x}_6$	(3)	0	0	0.2	−2	0	−1	1	2.4
2	Z	(0)	−1	0	0	$M + 0.5$	$1.6M - 1.1$	M	0	$-0.6M - 5.7$
	x_1	(1)	0	1	0	5	−1	0	0	3
	x_2	(2)	0	0	1	−5	3	0	0	9
	$\bar{x}_6$	(3)	0	0	0	−1	−0.6	−1	1	0.6

En la sección Worked Examples del sitio en internet del libro se proporciona **otro ejemplo** en el que al mismo problema se le aplica tanto el método de la gran *M* como el método de las dos fases.

Sin soluciones factibles

Hasta aquí, esta sección se ha ocupado más que nada del problema elemental de identificar la solución BF inicial cuando no se dispone de una obvia. Se explicó que se puede utilizar la técnica de variables artificiales para construir un problema artificial y obtener una solución BF inicial para este problema artificial. El método de la gran *M* o el de dos fases permiten al método símplex comenzar su recorrido hacia las soluciones BF y por último hacia la solución óptima del problema *real*.

No obstante, se debe estar consciente de que se puede presentar un obstáculo. Es posible que no exista una elección obvia para la solución BF inicial por la poderosa razón de que ¡no existan soluciones factibles! Cuando se construye una solución factible artificial, no hay nada que impida al método símplex proceder como siempre e incluso informar que encontró una supuesta solución óptima.

Por fortuna, la técnica de las variables artificiales proporciona algunas señales que indican que esto ha ocurrido:

> Si el problema original *no tiene soluciones factibles*, cualquier solución óptima que se obtenga con el método de la *M* o en la fase 1 del método de las dos fases lleva a una solución final que contiene al menos una variable artificial *mayor* que cero. De otra manera, *todas* son iguales a cero.

Para ilustrar lo que decimos, cambie la primera restricción del ejemplo de terapia de radiación (vea la figura 4.5) como sigue:

$$0.3x_1 + 0.1x_2 \leq 2.7 \quad \rightarrow \quad 0.3x_1 + 0.1x_2 \leq 1.8,$$

con lo que el problema ya no tiene soluciones factibles. Si se aplica el método de la gran *M* como antes (vea la tabla 4.12) se obtiene la tabla símplex que se muestra en la tabla 4.16. (La fase 1 del método de las dos fases conduce a la misma tabla símplex, excepto que cada expresión que involucra a *M* se reemplaza sólo por el factor multiplicativo.) Entonces, por lo común, el método de la gran *M* indicaría que la solución óptima es (3, 9, 0, 0, 0, 0.6). Sin embargo, en este caso,

ya que la variable artificial $\bar{x}_6 = 0.6 > 0$, el mensaje real es que el problema no tiene soluciones factibles.[16]

Variables que pueden ser negativas

En la mayor parte de los problemas prácticos, los valores negativos de las variables de decisión tienen un significado físico, por lo que es necesario incluir las restricciones de no negatividad en la formulación de los modelos de programación lineal. Sin embargo, esto no ocurre siempre. Como ejemplo, suponga que en el problema de la Wyndor Glass Co. el producto 1 ya está en producción y que la primera variable de decisión x_1 representa el *incremento* de la tasa de producción. En consecuencia, un valor negativo de x_1 indicaría que debe reducirse la fabricación del producto 1 en esa cantidad. Tal reducción podría ser deseable para permitir una tasa de producción más alta del nuevo producto 2, más rentable, con lo que se permitirían valores negativos de x_1 en el modelo.

Como el procedimiento para determinar la *variable básica saliente* requiere que todas las variables tengan restricción de no negatividad, cualquier problema que contenga variables que puedan adquirir valores negativos debe convertirse en un problema *equivalente* que emplee sólo variables no negativas antes de aplicar el método símplex. Por fortuna, se puede hacer esta conversión. La modificación que requiere cada variable depende de que tenga o no una cota inferior (negativa) sobre los valores permitidos. Se presentará cada uno de estos casos.

Variables con una cota sobre los valores negativos permitidos. Considere cualquier variable de decisión x_j que puede tener valores negativos, pero nada más aquellos que satisfacen una restricción de la forma

$$x_j \geq L_j,$$

donde L_j es una constante negativa. Esta restricción se puede convertir en una de no negatividad al cambiar de variables

$$x'_j = x_j - L_j, \quad \text{entonces} \quad x'_j \geq 0.$$

Así, $x'_j + L_j$ se sustituye por x_j en el modelo y la nueva variable de decisión x'_j no puede ser negativa. (Esta misma técnica se puede utilizar cuando L_j es *positiva* para convertir una restricción funcional $x_j \geq L_j$ en una restricción de no negatividad $x'_j \geq 0$.)

Para ejemplificar, suponga que la tasa de producción actual del producto 1 en el problema de la Wyndor Glass Co. es 10. Con la definición de x_1 que se acaba de dar, en este punto el modelo completo es el mismo que el que se dio en la sección 3.1, salvo que la restricción de no negatividad $x_1 \geq 0$ se sustituye por

$$x_1 \geq -10.$$

Para obtener el modelo equivalente que necesita el método símplex, la variable de decisión se redefinirá como la tasa de producción *total* del producto 1,

$$x'_j = x_1 + 10,$$

lo que produce los siguientes cambios en la función objetivo y las restricciones:

$$\boxed{\begin{array}{l} Z = 3x_1 + 5x_2 \\ x_1 \leq 4 \\ 2x_2 \leq 12 \\ 3x_1 + 2x_2 \leq 18 \\ x_1 \geq -10, \quad x_2 \geq 0 \end{array}} \rightarrow \boxed{\begin{array}{l} Z = 3(x'_1 - 10) + 5x_2 \\ x'_1 - 10 \leq 4 \\ 2x_2 \leq 12 \\ 3(x'_1 - 10) + 2x_2 \leq 18 \\ x'_1 - 10 \geq -10, \quad x_2 \geq 0 \end{array}} \rightarrow \boxed{\begin{array}{l} Z = -30 + 3x'_1 + 5x_2 \\ x'_1 \leq 14 \\ 2x_2 \leq 12 \\ 3x'_1 + 2x_2 \leq 48 \\ x'_1 \geq 0, \quad x_2 \geq 0 \end{array}}$$

[16] Se han desarrollado técnicas (y se han incorporado al software de programación lineal) para analizar qué ocasiona que un problema de programación lineal grande no tenga soluciones factibles, si en la formulación no se puede corregir ningún error. Por ejemplo, vea Chinneck, J. W., *Feasibility and Infeasibility in Optimization: Algorithms and Computational Methods*, Springer Science + Business Media, Nueva York, 2008.

Variables sin cota sobre los valores negativos permitidos. En caso de que x_j *no* tenga una cota inferior en el modelo formulado, se requiere un cambio distinto: x_j se sustituye en todo el modelo por la *diferencia* de dos nuevas variables *no negativas*,

$$x_j = x_j^+ - x_j^-, \qquad \text{donde } x_j^+ \geq 0,\ x_j^- \geq 0.$$

Como x_j^+ y x_j^- pueden tomar cualquier valor no negativo, la diferencia $x_j^+ - x_j^-$ puede tener *cualquier* valor (positivo o negativo), por lo que es una sustitución legítima de x_j en el modelo. Después de estas sustituciones, el método símplex puede proceder con variables que son no negativas.

Las nuevas variables x_j^+ y x_j^- tienen una interpretación sencilla. Como se explica en el siguiente párrafo, cada solución BF de la nueva forma del modelo tiene, necesariamente, la propiedad de que $x_j^+ = 0$ o $x_j^- = 0$ (o ambas). Por lo tanto, en la solución óptima que se obtuvo por el método símplex (una solución BF),

$$x_j^+ = \begin{cases} x_j & \text{si } x_j \geq 0, \\ 0 & \text{de otra manera;} \end{cases}$$

$$x_j^- = \begin{cases} |x_j| & \text{si } x_j \leq 0, \\ 0 & \text{de otra manera;} \end{cases}$$

de forma que x_j^+ representa la parte positiva de x_j y x_j^- su parte negativa (como lo sugieren los superíndices).

Por ejemplo, si $x_j = 10$, de las expresiones anteriores se obtiene $x_j^+ = 10$ y $x_j^- = 0$. Este mismo valor de $x_j = x_j^+ - x_j^- = 10$ ocurrirá también con valores grandes de x_j^+ y x_j^- tales que $x_j^+ = x_j^- + 10$. Cuando se grafican estos valores de x_j^+ y x_j^- en dos dimensiones se obtiene una recta con punto terminal en $x_j^+ = 10$, $x_j^- = 0$ para evitar violar las restricciones de no negatividad. Este punto es la única solución en un vértice sobre la recta. Por lo tanto, sólo este punto terminal puede ser parte de una solución FEV global o de la solución BF que involucra a todas las variables del modelo. Ello ilustra por qué necesariamente en cada solución BF se tiene que $x_j^+ = 0$ o $x_j^- = 0$ (o ambas).

Para ilustrar el uso de x_j^+ y x_j^- se regresará al ejemplo, dado en la página anterior en donde x_1 se vuelve a definir como el incremento sobre la tasa de producción actual de 10 del producto 1 en el problema de la Wyndor Glass Co.

Pero ahora suponga que la restricción $x_1 \geq -10$ no está incluida en el modelo original, ya que es claro que no influye en la solución óptima. (En algunos problemas, ciertas variables no necesitan tener cotas inferiores explícitas cuando las restricciones funcionales impiden valores menores.) Entonces, antes de aplicar el método símplex, x_1 se reemplaza por la diferencia,

$$x_1 = x_1^+ - x_1^-, \qquad \text{donde } x_1^+ \geq 0,\ x_1^- \geq 0,$$

como se muestra a continuación:

$$\boxed{\begin{array}{ll} \text{Maximizar} & Z = 3x_1 + 5x_2, \\ \text{sujeta a} & x_1 \leq 4 \\ & 2x_2 \leq 12 \\ & 3x_1 + 2x_2 \leq 18 \\ & x_2 \geq 0 \text{ (única)} \end{array}} \longrightarrow \boxed{\begin{array}{ll} \text{Maximizar} & Z = 3x_1^+ - 3x_1^- + 5x_2, \\ \text{sujeta a} & x_1^+ - x_1^- \leq 4 \\ & 2x_2 \leq 12 \\ & 3x_1^+ - 3x_1^- + 2x_2 \leq 18 \\ & x_1^+ \geq 0, \quad x_1^- \geq 0, \quad x_2 \geq 0 \end{array}}$$

Desde un punto de vista computacional, este enfoque tiene la desventaja de que el nuevo modelo equivalente tiene más variables que el modelo original. De hecho, si *ninguna* variable original tuviera restricción de cota inferior, el nuevo modelo tendría el *doble* de variables. Por fortuna, este enfoque se puede modificar en parte para que el número de variables aumente sólo en uno, sin importar cuántas variables originales tengan que sustituirse. Esta modificación se hace reemplazando cada variable de este tipo x_j por

$$x_j = x_j' - x'', \qquad \text{donde } x_j' \geq 0,\ x'' \geq 0,$$

donde x'' es la *misma* variable de toda *j* relevante. En este caso, la interpretación de x'' es que $-x''$ es el valor actual de la variable original negativa *más grande* (en términos de valor absoluto), por

TABLA 4.17 Análisis posóptimo para programación lineal

Tarea	Propósito	Técnica
Extracción de errores del modelo	Errores y debilidades del modelo	Reoptimización
Validación del modelo	Demostrar la validez del modelo final	Vea la sección 2.4
Decisión administrativa final sobre asignación de recursos (los valores b_i)	Efectuar una división apropiada de los recursos de la organización de actividades bajo estudio y otras actividades importantes	Precios sombra
Evaluación de las estimaciones de los parámetros del modelo	Determinar las estimaciones cruciales que pueden afectar la solución óptima de un estudio más amplio	Análisis de sensibilidad
Evaluación de trueques entre los parámetros del modelo	Determinar el mejor trueque	Programación lineal paramétrica

lo cual x_j' es la cantidad en la que x_j excede este valor. Con este recurso, el método símplex puede hacer que algunas x_j' adquieran valores mayores que cero aun cuando $x'' > 0$.

4.7 ANÁLISIS POSÓPTIMO

En las secciones 2.3, 2.4 y 2.5 se hizo hincapié en que el *análisis posóptimo*, el análisis que se hace *después* de obtener una solución óptima para la versión inicial del modelo, constituye una parte muy importante de casi todos los estudios de investigación de operaciones. En particular, este hecho es cierto en el caso de las aplicaciones comunes de programación lineal. Esta sección está dedicada a presentar el papel que juega el método símplex cuando se realiza este análisis.

En la tabla 4.17 se resumen los pasos que deben seguirse en un análisis posóptimo en estudios de programación lineal. En la última columna de ella se identifican algunas técnicas que emplea el método símplex. A continuación se ofrece una introducción breve a estas técnicas y los detalles se dejan para capítulos posteriores.

Reoptimización

Como se analizó en la sección 3.6, los modelos de programación lineal que surgen en la práctica casi siempre son muy grandes, con cientos o miles de restricciones funcionales y variables de decisión. En estos casos pueden ser de interés muchas variaciones del modelo básico que consideran diferentes escenarios. Por lo tanto, después de encontrar una solución óptima para una versión de un modelo de programación lineal, debe resolverse de nueva cuenta el problema (con frecuencia muchas veces) para una versión algo diferente de él. Casi siempre se deben resolver varias veces durante la etapa de extracción de errores del modelo (descrita en las secciones 2.3 y 2.4) y, por lo general, se hace lo mismo un gran número de veces durante las etapas de análisis posóptimo.

Una manera sencilla de hacerlo es aplicar el método símplex desde el principio a cada nueva versión del modelo, aunque cada corrida pueda requerir, en problemas grandes, cientos o miles de iteraciones. Sin embargo, una forma *mucho más eficiente* es la de *reoptimizar*. La reoptimización deduce los cambios que deben introducirse a la tabla símplex final (como se estudiará en las secciones 5.3 y 6.6). Esta tabla símplex revisada y la solución óptima del modelo anterior se usan como *tabla inicial* y *solución básica inicial* para resolver el nuevo modelo. Si esta solución es factible para el nuevo modelo, se aplica el método símplex en la forma usual, a partir de esta solución BF inicial. Si la solución no es factible, tal vez se pueda aplicar un algoritmo similar llamado *método símplex dual* (descrito en la sección 7.1) para encontrar la nueva solución óptima,[17] a partir de esta solución básica inicial.

La gran ventaja de la **técnica de reoptimización** sobre el hecho de volver a resolver el problema desde el principio, es que quizá la solución óptima del problema revisado esté *mucho más*

[17] El único requisito para usar el método símplex dual es que todavía se pase la *prueba de optimalidad* cuando se aplica al renglón 0 de la tabla símplex final *revisada*. Si no, en su lugar se puede usar otro algoritmo llamado *método primal-dual*.

cerca de la solución óptima anterior que de una solución BF inicial construida como siempre por el método símplex. En consecuencia, si se supone que las revisiones del modelo son moderadas, se requerirán sólo unas cuantas iteraciones para reoptimizar en lugar de cientos o tal vez miles que pueden realizarse al comenzar desde el principio. De hecho, las soluciones del modelo anterior y del revisado con frecuencia son las mismas, en cuyo caso, la técnica de reoptimización requiere sólo una aplicación de la prueba de optimalidad y *ninguna* iteración.

Precios sombra

Recuerde que, con frecuencia, los problemas de programación lineal se pueden interpretar como la asignación de recursos a las actividades. En particular, cuando las restricciones funcionales son de la forma $\leq$ las b_i (los lados derechos) se interpretan como las cantidades de los respectivos recursos disponibles para las actividades bajo estudio. En muchos casos puede haber dudas respecto de las cantidades que estarán disponibles. Si así es, los valores b_i que se usan en el modelo inicial (validado), en realidad pueden representar una *decisión inicial tentativa* del administrador sobre la cantidad de recursos de la organización que se asignarán a las actividades consideradas en el modelo y no a otras que él considere importantes. Desde esta perspectiva más amplia, algunos valores de b_i se pueden incrementar en un modelo revisado, pero sólo cuando se presenten razones poderosas sobre los beneficios que reportará esta revisión.

En consecuencia, la información sobre la contribución económica de los recursos a la medida de desempeño (Z) para el estudio en curso casi siempre será muy útil. El método símplex proporciona esta información en forma de *precios sombra* para los recursos respectivos.

> Los **precios sombra** del recurso i (denotados por y_i^*) miden el *valor marginal* de éste, es decir, la tasa a la que Z puede aumentar si se incrementa (un poco) la cantidad que se proporciona de este recurso (b_i).[18,19] El método símplex identifica este precio sombra como y_i^* = coeficiente de la i-ésima variable de holgura del renglón 0 de la tabla símplex final.

Como ejemplo, en el problema de la Wyndor Glass Co.,

Recurso i = capacidad de producción de la planta i ($i = 1, 2, 3$) que se proporciona para los dos nuevos productos bajo estudio,

b_i = horas del tiempo de producción por semana que se llevan a cabo en la planta i para estos nuevos productos.

La aportación de una cantidad sustancial de tiempo de producción para los nuevos productos requerirá un ajuste de los tiempos de producción de los productos actuales, por lo que la elección de los valores de b_i es una decisión administrativa difícil. La decisión inicial tentativa ha sido

$$b_1 = 4, \qquad b_2 = 12, \qquad b_3 = 18,$$

como se refleja en el modelo básico considerado en la sección 3.1 y en este capítulo. Sin embargo, ahora la administración desea evaluar el efecto de cambiar cualquiera de los valores de las b_i.

Los precios sombra de estos recursos proporcionan la información que necesita la administración. La tabla símplex final en la tabla 4.8 lleva a

$$y_1^* = 0 = \text{precio sombra del recurso 1},$$

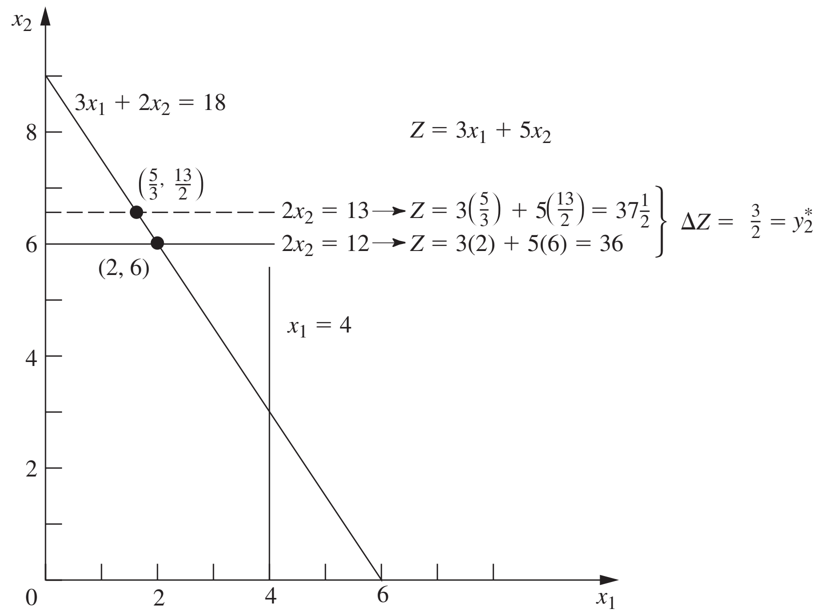
$$y_2^* = \frac{3}{2} = \text{precio sombra del recurso 2},$$

$$y_3^* = 1 = \text{precio sombra del recurso 3}.$$

Con sólo dos variables de decisión, estos números se pueden verificar en forma gráfica si se observa que un incremento individual de 1 en cualquier b_i aumentaría el valor de Z por y_i^*. Por ejemplo,

[18] El incremento de b_i debe ser suficientemente pequeño para que el conjunto actual de variables básicas siga siendo óptimo, ya que esta tasa (el valor marginal) cambia si el conjunto de variables básicas es otro.

[19] En caso de una restricción funcional de la forma $\geq$ o $=$ su precio sombra se define de nuevo como la tasa a la que puede aumentar Z al incrementar (un poco) el valor de b_i, aunque la interpretación usual de b_i ahora sería algo diferente a la cantidad de recursos disponibles.



■ **FIGURA 4.8**
La gráfica muestra que el precio sombra es $y_2^* = \frac{3}{2}$ del recurso 2 del problema de la Wyndor Glass Co. Los dos puntos son las soluciones óptimas de $b_2 = 12$ o $b_2 = 13$, y al insertar estas soluciones en la función objetivo se ve que un incremento de 1 en b_2 ocasiona un aumento de $y_2^* = \frac{3}{2}$ en *Z.*

en la figura 4.8 se muestra este incremento del recurso 2 cuando se aplica nuevamente el método gráfico que se presentó en la sección 3.1. La solución óptima, (2, 6) con $Z = 36$, cambia a $(\frac{5}{3}, \frac{13}{2})$ con $Z = 37\frac{1}{2}$ cuando b_2 aumenta en 1 (de 12 a 13), de manera que

$$y_2^* = \Delta Z = 37\frac{1}{2} - 36 = \frac{3}{2}.$$

Como *Z* se expresa en miles de dólares de ganancia semanal, $y_2^* = \frac{3}{2}$ indica que si se agregara una hora más de tiempo de producción a la semana en la planta 2 para estos nuevos productos, aumentaría la ganancia total en $1 500 semanales. ¿Debe hacerse esto? Depende de la ganancia marginal de otros productos que por el momento usan este tiempo de producción. Si existe un producto actual que contribuye con menos de $1 500 a la ganancia semanal por una hora de producción a la semana en la planta 2, entonces valdría la pena algún cambio en la asignación del tiempo de producción a los nuevos productos.

Esta historia continuará en la sección 6.7, donde el equipo de IO de la Wyndor utiliza los precios sombra como parte del *análisis de sensibilidad* del modelo.

En la figura 4.8 se demuestra que $y_2^* = \frac{3}{2}$ es la tasa a la que aumentaría *Z* si se incrementara un poco b_2 pero también demuestra el fenómeno general de que esta interpretación es válida nada más para aumentos pequeños de b_2. Si se aumenta a más de 18 unidades, la solución óptima se queda en el punto (0, 9) sin que *Z* aumente más. (En ese punto cambia el conjunto de variables básicas de la solución óptima y debe obtenerse una tabla símplex final con nuevos precios sombra, incluso $y_2^* = 0$.)

Ahora observe en la misma figura por qué $y_1^* = 0$. La restricción sobre el recurso 1, $x_1 \leq 4$, *no actúa en su frontera* en la solución óptima, (2, 6), ya que existe un *excedente* de este recurso. Por lo tanto, si b_1 adquiere un valor mayor que 4, no se obtendrá una nueva solución óptima con un valor mayor de *Z*.

Por el contrario, las restricciones sobre los recursos 2 y 3, $2x_2 \leq 12$ y $3x_1 + 2x_2 \leq 18$, son **restricciones satisfechas en sus fronteras** (son restricciones satisfechas en la igualdad en la solución óptima). Debido a que la disponibilidad limitada de estos recursos ($b_2 = 12$, $b_3 = 18$) *ata* a *Z* para que no pueda incrementarse, los precios sombra de estos recursos son *positivos*. Los economistas se refieren a este tipo de recursos como *bienes escasos*, mientras que los recursos disponibles en exceso (como el recurso 1) son *bienes libres* (recursos con precios sombra iguales a cero).

El tipo de información que proporcionan los precios sombra es valiosa para la administración cuando examina la posibilidad de reasignar recursos dentro de la organización. También resulta

muy útil cuando un aumento de b_i se puede lograr con sólo salir a comprar un poco más del recurso. Por ejemplo, suponga que Z representa *ganancias* y que las ganancias unitarias de las actividades (los valores de c_j) incluyen los costos (a precios normales) de todos los recursos que se consumen. Entonces, un precio sombra *positivo* de y_i^* del recurso significa que la ganancia total Z se puede aumentar en la cantidad y_i^* si se compra una unidad más de este recurso a su precio normal. Asimismo, si se tiene que pagar un precio *mayor al nominal* por la cantidad adicional del recurso y_i^* representará el precio *máximo* (cantidad adicional sobre el precio normal) que vale la pena pagar.[20]

La teoría de dualidad, que proporciona el fundamento teórico de los precios sombra, se describe en el capítulo 6.

Análisis de sensibilidad

Cuando se examinó el *supuesto de certidumbre* de la programación lineal al final de la sección 3.3, se hizo notar que los valores utilizados para los parámetros del modelo (las a_{ij}, b_i y c_j que se identifican en la tabla 3.3) casi siempre son sólo *estimaciones* de las cantidades cuyos verdaderos valores no se conocerán hasta que el estudio de programación lineal se lleve a la práctica en el futuro. El propósito principal del análisis de sensibilidad es identificar los **parámetros sensibles** (esto es, aquellos que no pueden cambiar sin modificar la solución óptima). Los parámetros sensibles son aquellos que será necesario controlar muy de cerca a medida que el estudio se ponga en práctica. Si se descubre que el valor verdadero de un parámetro sensible difiere de su valor estimado en el modelo, ello significa que la solución debe cambiar de inmediato.

¿Cómo se identifican los parámetros sensibles? En el caso de las b_i, acaba de verse que esta información está dada por los precios sombra que proporciona el método símplex. En particular, si $y_i^* > 0$, entonces la solución óptima cambia si b_i lo hace, por lo que b_i es un parámetro sensible. Sin embargo, $y_i^* = 0$ indica que la solución óptima no es sensible al menos a cambios pequeños en b_i. En consecuencia, si el valor que se usa para b_i es una estimación de la cantidad de recurso que se tendrá disponible (y no una decisión administrativa), los valores de b_i que se deben controlar con más cuidado son aquellos con precios sombra *positivos*, en especial los que tienen precios sombra *grandes*.

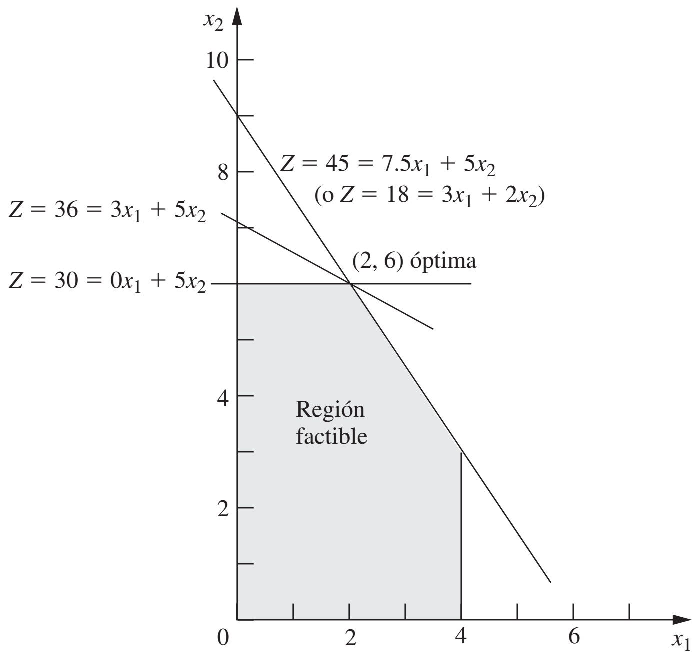
Cuando el problema tiene sólo dos variables, la sensibilidad de los distintos parámetros se puede analizar gráficamente. Por ejemplo, en la figura 4.9 se puede observar que $c_1 = 3$ puede cambiar a cualquier otro valor dentro del intervalo de 0 a 7.5 sin que la solución óptima (2, 6) cambie. (La razón es que cualquier valor de c_1 dentro de este intervalo mantiene la pendiente de $Z = c_1x_1 + 5x_2$ entre las pendientes de las líneas $2x_2 = 12$ y $3x_1 + 2x_2 = 18$.) De igual manera, si $c_2 = 5$ es el único parámetro que se cambia, puede tomar cualquier valor mayor que 2 sin que ello afecte la solución óptima. Ello nos indica que ni c_1 ni c_2 son parámetros sensibles. (El procedimiento llamado **Graphical Method and Sensitivity Analysis** que se incluye en el IOR Tutorial permite la realización de este tipo de análisis gráfico de una forma muy eficiente.)

La manera más fácil de analizar la sensibilidad de cada uno de los parámetros a_{ij} es verificar si su restricción correspondiente es satisfecha en su *frontera* en la solución óptima. Como $x_1 \leq 4$ *no* es una restricción satisfecha en su frontera, pues cualquier cambio suficientemente pequeño en sus coeficientes ($a_{11} = 1$, $a_{12} = 0$) no cambiará la solución óptima, así que éstos *no* son parámetros sensibles. Por otro lado, tanto $2x_2 \leq 12$ y $3x_1 + 2x_2 \leq 18$ son *restricciones satisfechas en su frontera*, por lo que al cambiar *cualquiera* de sus coeficientes ($a_{21} = 0$, $a_{22} = 2$, $a_{31} = 3$, $a_{32} = 2$) tendrá que cambiar la solución óptima, por lo cual éstos son parámetros sensibles.

Es común que se preste más atención al análisis de sensibilidad sobre los parámetros b_i y c_j que sobre los a_{ij}. En los problemas reales con cientos o miles de restricciones y variables, por lo general, el efecto que se produce al cambiar una a_{ij} es despreciable, mientras que el cambio de valor de una b_i o una c_j puede tener consecuencias notables. Aún más, en muchos casos, los valores de las a_{ij} están determinados por la tecnología que se usa (a veces se les da el nombre de *coeficientes tecnológicos*) por lo que puede que haya muy poca (o ninguna) incertidumbre respecto de sus valores finales. Esto resulta ventajoso puesto que en los problemas grandes existen muchos más parámetros a_{ij} que b_i y c_j.

---

[20] Si las ganancias unitarias *no* incluyen los costos de los recursos consumidos, entonces y_i^* representa el precio *total* unitario máximo que valdría la pena pagar para aumentar b_i.



■ **FIGURA 4.9**
La gráfica muestra el análisis de sensibilidad de c_1 y c_2 del problema de la Wyndor Glass Co. Comenzando con la recta de la función objetivo original [donde $c_1 = 3, c_2 = 5$, y la solución óptima es (2, 6)], las otras dos rectas muestran los extremos de cuánto puede cambiar la pendiente de la función objetivo sin que cambie la solución óptima (2, 6). Así, con $c_2 = 5$, el intervalo de valores permitido de c_1 es $0 \leq c_1 \leq 7.5$. Con $c_1 = 3$, el intervalo permisible de c_2 es $c_2 \geq 2$.

En problemas con más de dos (o tal vez tres) variables, no se puede analizar la sensibilidad de los parámetros en una gráfica como se hizo con el problema de la Wyndor Glass Co. Sin embargo, se puede extraer el mismo tipo de información del método símplex. Para obtenerla es preciso usar la *idea fundamental* que se describe en la sección 5.3 para deducir los cambios que se generan en la tabla símplex final como resultado de cambiar el valor de un parámetro en el modelo original. En las secciones 6.6 y 6.7 se describe y ejemplifica el resto del procedimiento.

Uso de Excel para generar información para el análisis de sensibilidad

Es común que el análisis de sensibilidad se incorpore en los paquetes de software basado en el método símplex. Por ejemplo, si se le pide, el Excel Solver genera información para el análisis de sensibilidad. Como se muestra en la figura 3.21, cuando Solver produce el mensaje de haber encontrado una solución, también da a la derecha una lista de tres informes que puede proporcionar. Si se selecciona el segundo (llamado "sensibilidad") después de resolver el problema de la Wyndor Glass Co., se obtendrá el *informe de sensibilidad* que se muestra en la figura 4.10. La tabla superior proporciona la información para el análisis de sensibilidad de las variables de decisión y sus coeficientes de la función objetivo. La tabla inferior hace lo mismo para las restricciones funcionales y los lados derechos.

■ **FIGURA 4.10**
Informe de sensibilidad proporcionado por Excel Solver sobre el problema de la Wyndor Glass Co.

Celdas cambiantes

Celda	Nombre	Valor final	Gradiente reducido	Coeficiente objetivo	Aumento permisible	Disminución permisible
$C$12	Lotes puertas	2	0	3 000	4 500	3 000
$D$12	Lotes ventanas	6	0	5 000	1E+30	3 000

Restricciones

Celda	Nombre	Valor final	Sombra precio	Restricción lado derecho	Aumento permisible	Disminución permisible
$E$7	Planta 1 Totales	2	0	4	1E+30	2
$E$8	Planta 2 Totales	12	1 500	12	6	6
$E$9	Planta 3 Totales	18	1 000	18	6	6

Observe primero la tabla superior de esta figura. La columna del "valor final" indica la solución óptima. La siguiente columna contiene los *costos reducidos* o gradiente reducido. (No se estudiarán estos costos ahora debido a que la información que proporcionan también se puede obtener del resto de la tabla de arriba.) Las siguientes tres columnas proporcionan la información necesaria para identificar el *intervalo permisible* para conservar la optimalidad de cada coeficiente c_j en la función objetivo.

> Para cualquier c_j, su **intervalo permisible** es el intervalo de valores de este coeficiente en el cual la solución óptima actual permanece óptima, siempre que no haya cambios en los otros coeficientes.

La columna del "coeficiente objetivo" proporciona el valor actual de cada coeficiente y las dos columnas siguientes el *aumento permisible* y la *disminución permisible* a partir de este valor para permanecer dentro del intervalo permisible. El modelo en hoja de cálculo (figura 3.22) expresa las ganancias por lote en unidades de *dólares*, mientras que la c_j de la versión algebraica del modelo de programación lineal utiliza unidades de *miles de dólares*, por lo que las cantidades que aparecen en estas tres columnas deben dividirse entre 1 000 para usar las mismas unidades que c_j. Por lo tanto,

$$\frac{3\,000 - 3\,000}{1\,000} \leq c_1 \leq \frac{3\,000 + 4\,500}{1\,000}, \qquad \text{entonces} \qquad 0 \leq c_1 \leq 7.5$$

es el intervalo permisible para c_1 en el cual la solución óptima actual permanecerá óptima (se supone que $c_2 = 5$), exactamente como se encontró con el método gráfico en la figura 4.9. De manera similar, como Excel usa 1E + 30 (10^{30}) para representar el infinito,

$$\frac{5\,000 - 3\,000}{1\,000} \leq c_2 \leq \frac{5\,000 + \infty}{1\,000}, \qquad \text{entonces} \qquad 2 \leq c_2$$

es el intervalo permisible para c_2 que conserva óptima la solución.

El hecho de que el aumento y la disminución permisibles sean mayores que cero para el coeficiente de ambas variables de decisión proporciona otra parte de la información útil, como se describe en seguida.

> Cuando la tabla superior del informe de análisis de sensibilidad generado por Excel Solver indica que tanto el aumento permisible como la disminución permisible son mayores que cero para todos los coeficientes objetivos, es una señal de que la solución óptima de la columna de "valor igual" es la única solución óptima. Por el contrario, algún aumento o disminución permisible igual a cero indica que existen soluciones múltiples. Al cambiar el coeficiente correspondiente en una cantidad muy pequeña a un valor mayor que cero y resolver de nuevo el problema se obtiene otra solución FEV óptima para el modelo original.

Ahora considere la tabla inferior de la figura 4.10 que se centra en el análisis de sensibilidad de las tres restricciones funcionales. La columna de "valor igual" proporciona el *valor del lado izquierdo* de cada restricción en la solución óptima. Las dos columnas siguientes dan los precios sombra y el valor actual de los lados derechos (b_i) de cada restricción. (Estos precios sombra del modelo en hoja de cálculo utilizan unidades de *dólares*, por lo cual deben dividirse entre 1 000 para usar las mismas unidades de *miles de dólares* que Z en la versión algebraica del modelo de programación lineal.) Cuando se cambia sólo un valor b_i las dos últimas columnas proporcionan el *aumento permisible* o la *disminución permisible* a fin de permanecer dentro del *intervalo permisible* para conservar su factibilidad.

> Para cualquier b_i, su **intervalo permisible** para conservar su factibilidad es el intervalo de valores del lado derecho en el cual la solución BF óptima actual (con valores ajustados[21] para las variables básicas) permanece factible, si se supone que no cambian los otros lados derechos. Una

[21] Debido a que los valores de las variables básicas se obtienen como la solución simultánea de un sistema de ecuaciones (las restricciones funcionales en la forma aumentada), al menos algunos de estos valores cambian si se modifica uno de los lados derechos. Sin embargo, los valores ajustados del conjunto actual de variables básicas todavía satisfacen las restricciones de no negatividad, de modo que todavía es factible, siempre y cuando el nuevo valor de este lado derecho permanezca dentro de su intervalo permisible para seguir factible. Si la solución básica ajustada sigue factible, también seguirá siendo óptima. Se profundizará sobre esta cuestión en la sección 6.7.

propiedad importante de este rango de valores es que el *precio sombra* actual de b_i permanece válido para evaluar el efecto sobre Z de cambiar b_i sólo mientras b_i permanezca dentro de este intervalo permisible.

Entonces, si se usa la tabla inferior de la figura 4.10, se combinan las dos últimas columnas con los valores actuales de los lados derechos para obtener los intervalos permisibles:

$$2 \le b_1$$
$$6 \le b_2 \le 18$$
$$12 \le b_3 \le 24.$$

El informe de sensibilidad generado por Excel Solver contiene la información característica del análisis de sensibilidad del software de programación lineal. En el apéndice 4.1 se verá que LINDO y LINGO proporcionan en esencia el mismo informe. MPL/CPLEX también lo hace cuando se pide en el cuadro de diálogo de "Solution File". Una vez más, esta información que se obtiene de manera algebraica también se puede derivar del análisis gráfico de este problema de dos variables (vea el problema 4.7-1). Por ejemplo, cuando b_2 se incrementa a más de 12 en la figura 4.8, la solución FEV óptima original en la intersección de las fronteras de las restricciones $2x_2 = b_2$ y $3x_1 + 2x_2 = 18$ seguirá factible (incluye $x_1 \ge 0$) sólo para $b_2 \le 18$.

En la sección Worked Examples del sitio en internet del libro se incluye **otro ejemplo** de aplicación del análisis de sensibilidad (en el que se usan tanto el análisis gráfico como un informe de sensibilidad). En la última parte del capítulo 6 se examinará este tipo de análisis con más detalle.

Programación lineal paramétrica

El análisis de sensibilidad requiere el cambio de un parámetro a la vez en el modelo original para examinar su efecto sobre la solución óptima. Por el contrario, la **programación lineal paramétrica** (o **programación paramétrica** en forma más corta) se refiere al estudio sistemático de los cambios en la solución óptima cuando cambia el valor de *muchos* parámetros *al mismo tiempo* dentro de un intervalo. Este estudio proporciona una extensión muy útil al análisis de sensibilidad; por ejemplo, se puede verificar el efecto de cambios simultáneos en parámetros "correlacionados", causados por factores exógenos tales como el estado de la economía. Sin embargo, una aplicación más importante es la investigación de los *trueques* entre los valores de los parámetros. Por ejemplo, si los valores de c_j representan la ganancia unitaria de las actividades respectivas, es posible aumentar el valor de alguna c_j a costa de disminuir el de otras mediante un intercambio apropiado de personal y equipo entre las actividades. De manera parecida, si los valores de b_i representan las cantidades disponibles de los respectivos recursos, es posible aumentar alguna b_i si se está de acuerdo en disminuir algunas otras. El análisis de este tipo de posibilidades se presenta e ilustra al final de la sección 6.7.

En algunas aplicaciones, el propósito del estudio es determinar el trueque más apropiado entre dos factores básicos como *costos* y *beneficios*. La forma usual de hacerlo es expresar uno de estos factores de la función objetivo (como minimizar el costo total) e incorporar el otro a las restricciones (por ejemplo, beneficio $\ge$ nivel mínimo aceptable), como se hizo en el problema de contaminación de la Nori & Leets Co., en la sección 3.4. La programación lineal paramétrica permite entonces la investigación sistemática de lo que ocurre cuando se cambia una decisión inicial tentativa sobre los trueques (como el nivel mínimo aceptable de los beneficios) a fin de mejorar un factor a costa de otro.

La técnica algorítmica para programación lineal paramétrica es una extensión natural del análisis de sensibilidad, por lo que también está basada en el método símplex. El procedimiento se describe en la sección 7.2.

4.8 USO DE COMPUTADORA

Si la computación electrónica no se hubiera inventado, no se oiría hablar de programación lineal ni del método símplex. Aunque es posible aplicar el método símplex a mano (probablemente con la ayuda de una calculadora) para resolver problemas muy pequeños de programación lineal, los cálculos necesarios son demasiado tediosos para llevarlos a cabo de manera rutinaria. Sin embargo, el método símplex es un algoritmo muy adecuado para su ejecución en computadora. La revolu-

ción de las computadoras ha hecho posible la amplia aplicación de la programación lineal en las últimas décadas.

Implantación del método símplex

Existe una amplia disponibilidad de programas del método símplex en esencia para casi todos los sistemas de cómputo modernos. Por lo general, estos programas son parte de paquetes de software complejos de programación matemática que incluyen muchos procedimientos descritos en los capítulos subsecuentes (incluso los referentes a análisis posóptimo).

Estos programas de computadora no siempre siguen la forma tabular o la algebraica del método símplex que se presentaron en las secciones 4.3 y 4.4. Estas formas se pueden simplificar de manera significativa al llevarlas a la computadora. Por lo tanto, los programas usan una *forma matricial* (que suele llamarse *método símplex revisado*) muy adecuada para computación. Esta forma obtiene los mismos resultados que las formas tabular y algebraica, pero calcula y almacena sólo los números que necesita para la iteración actual, y después guarda los datos esenciales de manera más compacta. En las secciones 5.2 y 5.4 se describe este método.

El método símplex se usa en forma rutinaria para resolver problemas de programación lineal sorprendentemente grandes. Por ejemplo, algunas computadoras personales poderosas (en especial las estaciones de trabajo) pueden resolver problemas con cientos de miles o, inclusive, de millones de restricciones funcionales y un número mayor de variables de decisión. En ocasiones, problemas resueltos con éxito tienen inclusive decenas de millones de restricciones funcionales y variables de decisión.[22] Para ciertos *tipos especiales* de problemas de programación lineal (como problemas de transporte, asignación y flujo de costo mínimo que se describirán en secciones posteriores), ahora pueden resolverse problemas de tamaño mucho mayor con versiones *especializadas* del método símplex.

Varios factores afectan el tiempo que tarda el método símplex general para resolver un problema de programación lineal. El más importante es el *número de restricciones funcionales ordinarias*. De hecho, el tiempo de cálculo tiende a ser, a grandes rasgos, proporcional al cubo de este número, por lo que al duplicarlo el tiempo puede quedar multiplicado por un factor aproximado a 8. Por el contrario, el número de variables es un factor de importancia relativamente menor.[23] Así, aunque se dupliquen, tal vez ni siquiera se duplique el tiempo de cálculos. Un tercer factor con alguna importancia es la *densidad* de la tabla de coeficientes de las restricciones (es decir, la *proporción* de coeficientes *distintos* de cero) ya que afecta el tiempo de cálculo *por iteración*. (En problemas grandes que se encuentran en la práctica, es común que la densidad esté por debajo de 5% e incluso sea menor a 1%, "dispersión" que tiende a acelerar el método símplex de manera considerable.) Una regla empírica común para estimar el *número de iteraciones* es que tiende a ser el doble del número de restricciones funcionales.

En problemas de programación lineal grandes es inevitable que al inicio se cometan algunos errores y se tomen decisiones equivocadas cuando se formula el modelo y se lo introduce en la computadora. En consecuencia, como se estudió en la sección 2.4, se necesita un proceso exhaustivo de pruebas y refinamiento (*validación del modelo*). El producto terminal usual no es un modelo estático que el método símplex resuelve de una sola vez. Por el contrario, el equipo de IO y casi siempre la administración toman en cuenta una larga serie de variaciones sobre el modelo básico (en ocasiones, miles de variaciones) para examinar diferentes escenarios como parte del análisis posóptimo. Este proceso completo se acelera en gran medida cuando se puede llevar a cabo de manera *interactiva* en una *computadora personal*. Con la ayuda de los lenguajes de modelado matemático y de los avances en la tecnología de computadoras este procedimiento es, cada vez más, una práctica común.

Hasta mediados de la década de 1980, los problemas de programación lineal se resolvían casi de manera exclusiva en *computadoras grandes*. Un interesante desarrollo reciente es la explosiva

[22] No intente desarrollar esta tarea en casa. Los problemas de este tipo requieren un sistema de programación lineal complejo que usa las técnicas modernas que aprovechan la proporción de coeficientes en la matriz y otras técnicas especiales (por ejemplo, *técnicas de quiebre* para encontrar con rapidez una solución BF inicial avanzada). Cuando se resuelven problemas en forma periódica después de alguna actualización menor de los datos, con frecuencia se ahorra mucho tiempo cuando se utiliza (o se modifica) la última solución óptima como solución inicial básica factible de la nueva corrida.

[23] Esta afirmación supone que se utiliza el método símplex revisado descrito en las secciones 5.2 y 5.4.

ampliación de la capacidad de ejecutar programas de programación lineal en las computadoras personales, las microcomputadoras y las estaciones de trabajo. Éstas, que incluyen algunas con capacidad de procesamiento en paralelo, en la actualidad se emplean en lugar de las computadoras grandes para resolver modelos masivos. Las computadoras personales más rápidas no se quedan atrás aunque la solución de modelos extensos suele requerir memoria adicional.

Software de programación lineal descrito en este libro

Hoy se dispone de un gran número de paquetes de software para programación lineal y sus extensiones, que satisfacen distintas necesidades. Uno de los más avanzados de este tipo es **Express-MP**, un producto de Dash Optimization (que se unió con Fair Isaac). Otro paquete que es considerado muy poderoso para la resolución de problemas extensos es **CPLEX**, un producto de ILOG, Inc., con oficinas en Silicon Valley. Desde 1988, CPLEX ha marcado el camino de la solución de problemas de programación lineal cada vez más grandes. Los extensos esfuerzos de investigación y desarrollo han permitido una serie de actualizaciones con incrementos drásticos del nivel de eficiencia. CPLEX 11, que se liberó en 2007, brinda otra mejora con una magnitud importante. Este software ha resuelto con éxito problemas de programación lineal reales surgidos en la industria en decenas de millones de restricciones funcionales y variables de decisión! Con frecuencia, CPLEX utiliza el método símplex y sus variantes (como el método símplex dual que se presentó en la sección 7.1) para resolver estos problemas masivos. Además del método símplex, CPLEX cuenta con otras herramientas poderosas para resolver problemas de programación lineal. Una de ellas es un algoritmo muy rápido que aplica el *enfoque de punto interior* que se introduce en la sección 4.9. Este algoritmo puede resolver algunos problemas de programación lineal enormes que el método símplex no puede (y viceversa). Otra presentación es el *método símplex de redes* (descrito en la sección 9.7), que puede resolver tipos especiales de problemas de programación lineal aún más grandes. CPLEX 11 también va más allá de la programación lineal e incluye algoritmos modernos para *programación entera* (capítulo 11) y *programación cuadrática* (sección 12.7), así como *programación cuadrática entera*.

Es posible predecir que estos avances importantes de los paquetes de software para optimización, como CPLEX, continuarán en el futuro. La continua mejora de la velocidad de las computadoras también permite anticipar un acelerado aumento de la velocidad de los futuros paquetes de software.

Como a menudo se utiliza para resolver problemas realmente grandes, lo normal es que CPLEX se use junto con un *lenguaje de modelado* de programación matemática. Como se describió en la sección 3.7, los lenguajes de modelado están diseñados para formular con eficiencia modelos de programación lineal grandes (y modelos relacionados) de manera compacta, después de lo cual se corre un solucionador para resolver el modelo. Varios lenguajes de modelado sobresalientes trabajan con CPLEX como solucionadores. ILOG introdujo también su propio lenguaje de modelado, llamado *Lenguaje de Programación para Optimización* (*OPL*) que se puede usar con CPLEX para formar el *Sistema de Desarrollo OPL-CPLEX.* (Se puede encontrar una versión de prueba de ese producto en la página de internet de ILOG, www.ilog.com.)

Como se mencionó en la sección 3.7, la versión para estudiantes de CPLEX se incluye en el OR Courseware como el solucionador del lenguaje de modelado MPL.

La versión del estudiante de MPL que contiene el OR Courseware también incluye dos solucionadores que representan una alternativa de CPLEX para resolver problemas de programación lineal y problemas de programación con enteros (temas que se analizan en el capítulo 11). Uno de ellos es **CoinMP**, un solucionador de software que puede resolver problemas más complejos que la versión del estudiante de CPLEX (el cual está limitado a 300 restricciones y variables). El otro es LINDO.

LINDO (iniciales de Linear INteractive and Discrete Optimizer) tiene una historia más antigua que CPLEX en el campo de las aplicaciones de la programación lineal y sus extensiones. La interfaz fácil de usar de LINDO se encuentra disponible como un subconjunto del paquete de modelado de optimización **LINGO** de LINDO Systems, www.lindo.com. En parte, la duradera popularidad de LINDO se debe a su sencillez de manejo. Para problemas relativamente pequeños (tamaño libro de texto), el modelo se puede introducir y resolver de manera bastante intuitiva, por lo que se trata de una herramienta útil para los estudiantes. Aunque su uso es muy sencillo en modelos pequeños, LINDO/LINGO puede resolver también modelos de gran tamaño, por ejemplo, la

versión más grande de este software ha sido capaz de resolver problemas reales con 4 millones de variables y 2 millones de restricciones.

En el OR Courseware que se proporciona en el sitio de internet del libro se incluye la versión para estudiantes de LINGO/LINDO junto con un tutorial extenso. El apéndice 4.1 proporciona una breve introducción. Además, el software contiene una gran cantidad de ayuda en línea. El OR Courseware también cuenta con formulaciones LINGO/LINDO para los ejemplos principales que se utilizan en este libro.

Los solucionadores basados en las hojas de cálculo son cada vez mejor recibidos para programación lineal y sus extensiones. Entre los líderes se encuentran los solucionadores producidos por Frontline Systems para Microsoft Excel y otros paquetes de hojas de cálculo. Además del solucionador básico que incluyen estos paquetes, se dispone de otros más poderosos como el *Premium Solver*. Debido al amplio uso actual del software de hojas de cálculo como Microsoft Excel, estos solucionadores permiten que un creciente número de personas conozcan el potencial de la programación lineal. Para problemas tipo libros de texto (y algunos bastante más grandes), las hojas de cálculo proporcionan una manera conveniente de formular y resolver modelos de programación lineal, como se describe en la sección 3.5. Los solucionadores más poderosos pueden resolver modelos bastante grandes con varios miles de variables de decisión. Sin embargo, cuando la hoja de cálculo crece hasta un tamaño difícil de manejar, un buen lenguaje de modelado y su solucionador pueden proporcionar un enfoque más eficiente para formular y resolver el modelo.

Las hojas de cálculo proporcionan una excelente herramienta de comunicación, en especial si se trata de administradores típicos que se sienten a gusto con este formato pero no con las formulaciones algebraicas de IO. Por esta razón, es común que los paquetes de optimización y los lenguajes de modelado puedan importar y exportar datos y resultados al formato de una hoja de cálculo. Por ejemplo, en la actualidad, el lenguaje de modelado MPL incluye una mejora (llamada *OptiMax 2000 Component Library*) que permite crear la sensación de un modelo en hoja de cálculo para el usuario del modelo pero que usa MPL para formularlo de manera muy eficiente. (La versión para estudiantes de OptiMax 2000 se incluye en el OR Courseware.)

El **Premium Solver for Education** es uno de los complementos de Excel incluido en el CD-ROM. Se puede instalar para obtener un desempeño mucho mejor que con el Excel Solver estándar.

En consecuencia, todo el software, los tutoriales y los ejemplos que se incluyen en el sitio de internet del libro proporcionan varias opciones atractivas de paquetes para programación lineal.

Opciones de software disponibles para programación lineal

1. Ejemplos de demostración (en OR Tutor) y procedimientos interactivos y automáticos en el IOR Tutorial para un aprendizaje eficiente del método símplex.
2. Excel y el Premium Solver para formular y resolver modelos de programación lineal en una hoja de cálculo.
3. MPL/CPLEX para formular y solucionar de manera eficiente modelos grandes de programación lineal.
4. LINGO y su solucionador (compartido con LINDO) que proporciona una manera alternativa de formular y resolver modelos grandes de programación lineal con eficiencia.
5. LINDO para formular y resolver modelos de programación lineal en forma directa.

Quizá el instructor del curso especifique cuál software es conveniente utilizar. Cualquiera que sea la elección se adquirirá experiencia con el tipo de software actual que usan los profesionales de IO.

4.9 ENFOQUE DE PUNTO INTERIOR PARA RESOLVER PROBLEMAS DE PROGRAMACIÓN LINEAL

El desarrollo de más trascendencia durante la década de 1980 en investigación de operaciones fue el descubrimiento del enfoque de punto interior para resolver problemas de programación lineal. Este descubrimiento se efectuó en 1984, cuando un joven matemático de AT&T Bell Laboratories, Narendra Karmarkar, anunció el desarrollo de un nuevo algoritmo para resolver problemas de programación lineal con este enfoque. Aunque el algoritmo experimentó sólo una mezcla de

aceptación y rechazo cuando comenzó a competir con el método símplex, el concepto clave de solución que describía parecía tener un gran potencial para resolver problemas de programación lineal *enormes* más allá del alcance del método símplex. Muchos investigadores de alto nivel trabajaron en modificaciones del algoritmo de Karmarkar para aprovechar todo su potencial. Se han logrado grandes avances (y continúan) y se han desarrollado varios algoritmos poderosos que usan el enfoque de punto interior. En la actualidad, los paquetes de software más eficaces diseñados para resolver problemas muy grandes (como CPLEX) incluyen al menos un algoritmo que emplea este enfoque junto con el método símplex y sus variantes. La investigación sobre estos algoritmos continúa y su implantación en la computadora sigue mejorando. Estos avances han renovado la investigación sobre el método símplex lo mismo que sobre su implantación en computadoras. A la vez, se mantiene la competencia por la supremacía entre los dos enfoques para resolver problemas muy grandes.

A continuación se estudiará la idea clave en la que se apoya el algoritmo de Karmarkar y sus variantes subsecuentes que aplican el enfoque de punto interior.

El concepto de solución clave

Aunque es bastante diferente del método símplex, el algoritmo de Karmarkar comparte algunas de sus características. En principio, es un algoritmo *iterativo*. Comienza por identificar una *solución de prueba* factible. En cada iteración, se mueve de una solución de prueba actual a una mejor solución de prueba en la región factible. Después continúa este proceso hasta llegar a una solución de prueba que es (en esencia) óptima.

La gran diferencia se encuentra en la naturaleza de estas soluciones de prueba. En el caso del método símplex, éstas son *soluciones FEV* (o soluciones BF en el problema aumentado), de manera que todos los movimientos se hacen por las aristas sobre la *frontera* de la región factible. De acuerdo con el algoritmo de Karmarkar, las soluciones de prueba son **puntos interiores**, es decir, puntos *dentro* de la frontera de la región factible. Ésta es la razón por la que se hace referencia al algoritmo de Karmarkar y sus variantes como **algoritmos de punto interior**.

Sin embargo, debido a que se obtuvo una patente de una versión anterior de un algoritmo de punto interior, con frecuencia a este algoritmo también se le denomina **algoritmo de barrera** (o *método de barrera*). El término *barrera* se utiliza porque desde la perspectiva de una búsqueda cuyas soluciones de prueba son *puntos interiores*, cada frontera de restricción es tratada como una barrera. En la actualidad, la mayoría de los paquetes de software para optimización utilizan la terminología de la barrera cuando se refieren a su opción de solucionador basado en el enfoque del punto interior. Tanto CPLEX como LINDO API incluyen un "algoritmo de barrera" que se puede emplear para resolver problemas de programación lineal o cuadrática (estos últimos se explican en la sección 12.7).

Para ilustrar este enfoque, en la figura 4.11 se muestra la trayectoria seguida por el algoritmo de punto interior en el OR Courseware cuando se aplica al problema de la Wyndor Glass Co. que comienza en la solución de prueba inicial (1, 2). Observe que todas las soluciones de prueba (puntos) que aparecen en esta trayectoria se encuentran dentro de la frontera de la región factible a medida que la trayectoria se acerca a la solución óptima (2, 6). (Todas las soluciones de prueba subsecuentes que no se muestran están también dentro de la frontera de la región factible.) Compare esta trayectoria con la que se sigue en el método símplex alrededor de la frontera de (0, 0) a (0, 6) a (2, 6).

En la tabla 4.18 se muestra una salida real del OR Courseware para este problema.[24] (Inténtelo.) Observe cómo las soluciones de prueba sucesivas se acercan cada vez más a la solución óptima, pero en realidad nunca llegan. Sin embargo, la desviación se vuelve infinitesimalmente pequeña y, para propósitos prácticos, la solución de prueba final puede tomarse como la solución óptima. (En la sección Worked Examples del sitio en internet del libro se muestra el resultado del IOR Tutorial para **otro ejemplo**.)

En la sección 7.4 se presentan los detalles del algoritmo de punto interior específico que se desarrolló en el IOR Tutorial.

[24] La rutina se llama *Solve Automatically by the Interior-Point Algorithm* (solución automática por el algoritmo de punto interior). El menú de "option" da dos alternativas para cierto parámetro del algoritmo (definido en la sección 7.4). La elección usada aquí es el valor establecido de $\alpha = 0.5$.

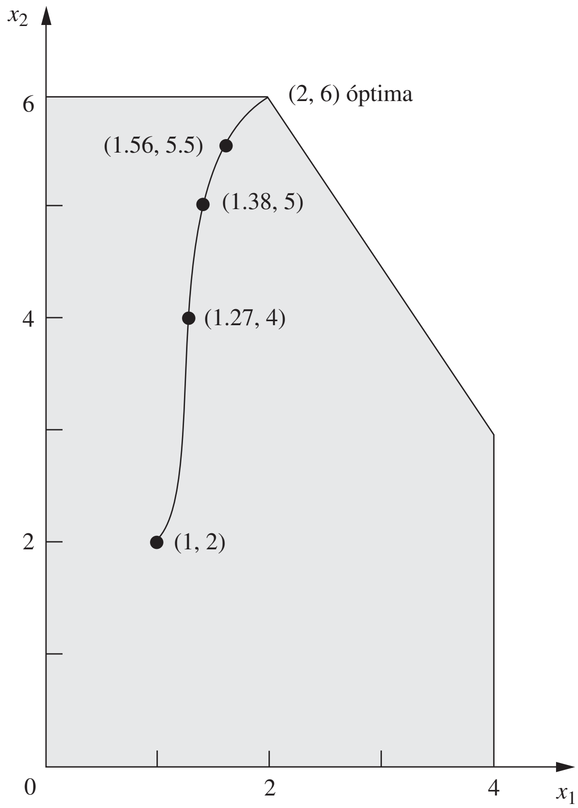


FIGURA 4.11
La curva de (1, 2) a (2, 6) muestra una trayectoria típica seguida por un algoritmo de punto interior, a través de la parte *interna* de la región factible del problema de la Wyndor Glass Co.

Comparación con el método símplex

Una manera significativa de comparar los algoritmos de punto interior con el método símplex es examinar sus propiedades teóricas en cuanto a complejidad computacional. Karmarkar demostró que la versión original de su algoritmo es un **algoritmo de tiempo polinomial**; es decir, el tiempo que se requiere para resolver *cualquier* problema de programación lineal se puede acotar por arriba mediante una función polinomial del tamaño del problema. Se han construido contraejemplos muy

TABLA 4.18 Salida del algoritmo de punto interior en OR Courseware del problema de la Wyndor Glass Co.

Iteración	x_1	x_2	Z
0	1	2	13
1	1.27298	4	23.8189
2	1.37744	5	29.1323
3	1.56291	5.5	32.1887
4	1.80268	5.71816	33.9989
5	1.92134	5.82908	34.9094
6	1.96639	5.90595	35.429
7	1.98385	5.95199	35.7115
8	1.99197	5.97594	35.8556
9	1.99599	5.98796	35.9278
10	1.99799	5.99398	35.9639
11	1.999	5.99699	35.9819
12	1.9995	5.9985	35.991
13	1.99975	5.99925	35.9955
14	1.99987	5.99962	35.9977
15	1.99994	5.99981	35.9989

rebuscados para demostrar que el método símplex no posee esta propiedad y que es un **algoritmo de tiempo exponencial** (esto es, el tiempo requerido sólo puede ser acotado por arriba mediante una función exponencial del tamaño del problema). La diferencia de *desempeño en el peor de los casos* es considerable. No obstante, esto no dice nada sobre la comparación de desempeño promedio en problemas reales, que es un asunto de mayor importancia.

Los dos factores básicos que determinan el desempeño de un algoritmo ante un problema real son el *tiempo promedio de computadora por iteración* y el *número de iteraciones*. Las siguientes comparaciones tratan sobre estos factores.

Los algoritmos de punto interior son mucho más complicados que el método símplex. Se requieren muchos más cálculos en cada iteración para encontrar la siguiente solución de prueba. Por lo tanto, el tiempo de computadora por iteración de un algoritmo de punto interior es muchas veces mayor que el que emplea el método símplex.

En el caso de problemas claramente pequeños, el número de iteraciones que requiere el algoritmo de punto interior y el método símplex tienden a ser comparables. Por ejemplo, en un problema con 10 restricciones funcionales, 20 iteraciones sería lo normal para los dos tipos de algoritmos. En consecuencia, en problemas de tamaño similar, el tiempo total de computadora para desarrollar un algoritmo de punto interior tenderá a ser mucho mayor que si se utiliza el método símplex.

Por otro lado, una ventaja de los algoritmos de punto interior es que los problemas grandes no requieren muchas más iteraciones que los problemas pequeños. Por ejemplo, un problema con 10 000 restricciones funcionales tal vez requiera menos de 100 iteraciones. Aun si se considera el tiempo sustancial de computadora necesario por iteración para un problema de este tamaño, un número de iteraciones tan pequeño hace que el problema sea muy manejable. Por el contrario, el método símplex puede requerir 20 000 iteraciones, lo que genera el peligro de no terminar en un tiempo razonable. En consecuencia, es muy probable que los algoritmos de punto interior sean más rápidos que el método símplex para problemas tan grandes.

La razón de esta gran diferencia en el número de iteraciones para problemas grandes es la diferencia entre las trayectorias que se siguen. En cada iteración el método símplex se mueve de la solución FEV actual a una solución FEV adyacente por una arista de la frontera de la región factible. Los problemas grandes tienen cantidades astronómicas de soluciones FEV. La trayectoria desde la solución FEV inicial hasta una solución óptima puede dar muchas vueltas por la frontera, y dar numerosos pasos antes de llegar a cada solución FEV adyacente siguiente. En contraste, un algoritmo de punto interior evita este engorroso deambular pues avanza por el interior de la región factible hacia la solución óptima. Al agregar restricciones funcionales se agregan aristas a la frontera de la región factible, lo cual tiene muy poco efecto sobre el número de soluciones de prueba que se necesitan en la trayectoria a través del interior. Esta característica posibilita que los algoritmos de punto interior resuelvan problemas con un número muy grande de restricciones funcionales.

Una última comparación importante se refiere a la capacidad para realizar los distintos tipos de análisis posóptimo descritos en la sección 4.7. El método símplex y sus extensiones son muy adecuados y su uso es muy extenso para este tipo de análisis. Por ejemplo, el producto de ILOG llamado *Optimization Decision Manager* hace un uso extenso del método símplex en CPLEX para llevar a cabo una gran variedad de tareas del análisis de posoptimización de una manera apropiada. Desafortunadamente, en la actualidad el enfoque de punto interior tiene una capacidad muy limitada en esta área.[25] Dada la gran importancia del análisis posóptimo, ésta es una falla crucial de los algoritmos de punto interior. Sin embargo, se establecerá una forma en que el método símplex se puede combinar con el enfoque de punto interior para salvar este obstáculo.

Papeles complementarios del método símplex y el enfoque de punto interior

La investigación que se realiza en la actualidad proporciona mejoras sustanciales en la implantación en computadora del método símplex (incluidas sus variantes) y el algoritmo de punto interior. Por lo tanto, cualquier predicción sobre su papel futuro es riesgosa. De todas maneras se presentará un resumen de la predicción actual acerca de sus papeles complementarios.

[25] Sin embargo, la investigación dirigida a aumentar esta capacidad continúa en avance. Por ejemplo, vea E. A. Yildirim y M. J. Todd, "Sensitivity Analysis in Linear Programming and Semidefinite Programming Using Interior-Point Methods", en *Mathematical Programming*, Series A, **90**(2): 229-261, abril de 2001.

El método símplex (y sus variantes) todavía es el algoritmo estándar para el uso rutinario de programación lineal. Es el algoritmo más eficiente para resolver problemas con menos de, digamos, 10 000 restricciones funcionales. También es el más eficiente para abordar algunos (pero no todos) de los problemas con hasta, digamos, 100 000 restricciones funcionales y casi un número ilimitado de variables de decisión, por lo que la mayor parte de los usuarios aún lo utilizan para este tipo de problemas. Sin embargo, cuando el número de restricciones funcionales aumenta, cada vez es más probable que el enfoque de punto interior sea el más eficiente, por lo cual en la actualidad se lo emplea con mayor frecuencia. Cuando el tamaño llega a cientos de miles o inclusive millones de restricciones funcionales, el enfoque de punto interior puede ser el único capaz de resolver el problema. Aun así, no siempre ocurre de este modo. Como se mencionó en la sección anterior, el software moderno ha tenido éxito en el uso del método símplex y sus variantes para resolver algunos problemas masivos con millones e incluso decenas de millones de restricciones y variables de decisión.

Estas generalizaciones sobre la comparación entre el enfoque de punto interior y el método símplex para distintos tamaños de problemas no son acertadas en todos los casos. El software y el equipo de cómputo que se usan tienen un efecto importante. El *tipo específico* de problema de programación lineal que se desea resolver afecta de manera notable la comparación. Al pasar el tiempo se sabrá más sobre cómo identificar los tipos de problemas que son más adecuados para cada tipo de algoritmo.

Uno de los productos secundarios del enfoque de punto interior ha sido la renovación importante de los esfuerzos dedicados a mejorar la eficiencia de los programas de computadora del método símplex. Se han logrado progresos impresionantes en los últimos años y habrá más por venir. Al mismo tiempo, la investigación y desarrollo en marcha sobre el enfoque de punto interior aumentará todavía más su poder, y quizá a un paso más acelerado que en el caso del método símplex.

El mejoramiento de la tecnología de las computadoras, como el procesamiento masivo en paralelo (un gran número de unidades de computadora que operan en paralelo en distintas partes del mismo problema), también significará un aumento sustancial del tamaño del problema que cualquiera de los dos tipos de algoritmos pueda resolver. Sin embargo, por el momento parece que los enfoques de punto interior tienen mucho mayor potencial que el método símplex para aprovechar el procesamiento en paralelo.

Como ya se dijo, una desventaja grave del enfoque de punto interior es su limitada capacidad para realizar un análisis posóptimo. Para compensar esta falla, los investigadores han intentado desarrollar procedimientos para cambiar al método símplex cuando el algoritmo de punto interior termina. Recuerde que las soluciones pruebas obtenidas por éste se acercan cada vez más a una solución óptima (la mejor solución FEV), pero nunca llegan a ella. Por esta razón, un procedimiento para cambiar de enfoque requiere identificar la solución FEV (o la solución BF del problema aumentado) que es muy cercana a la solución de prueba final.

Por ejemplo, si se observa la figura 4.11, es sencillo ver que la solución de prueba final de la tabla 4.18 es muy cercana a (2, 6). Desafortunadamente, en problemas con miles de variables de decisión (en los que no se dispone de una gráfica), la identificación de una solución FEV (o BF) cercana es una tarea difícil y lenta. No obstante, se ha progresado bastante en el desarrollo de procedimientos para hacerlo. Por ejemplo, la versión profesional completa de CPLEX incluye un *algoritmo superior* que convierte la solución obtenida mediante su "algoritmo de barrera" en una solución BF.

Una vez que se encuentra la solución BF cercana, se aplica la prueba de optimalidad del método símplex para verificar si se trata de la solución BF óptima. Si no lo es, se llevan a cabo algunas iteraciones del método símplex para moverla hacia esa dirección. En general, sólo se necesitan unas cuantas iteraciones (tal vez una sola) ya que el algoritmo de punto interior se acerca de manera estrecha a una solución óptima. Estas iteraciones deben poderse realizar bastante rápido, aun en problemas demasiado grandes para resolverlos desde el principio. Después de obtener realmente una solución óptima, se aplican el método símplex y sus extensiones para ayudar a realizar el análisis posóptimo.

Debido a las dificultades que implica la aplicación de un procedimiento de cambio (que incluye el tiempo adicional de computadora), algunos profesionales prefieren usar sólo el método símplex desde el principio. Ésta parece una buena opción cuando sólo en algunas ocasiones se encuentran problemas suficientemente grandes como para que un algoritmo de punto interior sea

un poco más rápido (antes de hacer el cambio) que el método símplex. Esta modesta aceleración no justificaría ni el tiempo adicional de computadora para el procedimiento de cambio ni el alto costo de adquirir (y aprender a usar) el software basado en el enfoque de punto interior. Sin embargo, para las organizaciones que con frecuencia deben manejar problemas de programación lineal extremadamente grandes, la adquisición del nuevo software de este tipo (con el procedimiento de cambio) tal vez valga la pena. Cuando se trata de problemas enormes, la única forma de solución disponible pueden ser estos paquetes.

Algunas veces, las aplicaciones de los modelos de programación lineal muy grandes generan ahorros de millones de dólares. Sólo una de ellas puede pagar varias veces el costo de los paquetes de software basados en el enfoque de punto interior y el cambio al método símplex al final.

4.10 CONCLUSIONES

El método símplex es un algoritmo eficiente y confiable para resolver problemas de programación lineal. También proporciona la base para llevar a cabo, en forma muy eficiente, las distintas etapas del análisis posóptimo.

Aunque tiene una interpretación geométrica útil, el método símplex es un procedimiento algebraico. En cada iteración se mueve de la solución BF actual a una adyacente mejor mediante la elección de la variable básica entrante y de la saliente; después recurre a la eliminación de Gauss para resolver el sistema de ecuaciones lineales. Cuando la solución actual no tiene una solución BF adyacente que sea mejor, la solución actual es óptima y el algoritmo se detiene.

Se presentó la forma algebraica completa del método símplex para establecer su lógica y se llevó el método a una forma tabular más conveniente. Para preparar el inicio del método símplex, algunas veces es necesario usar variables artificiales para obtener una solución básica factible inicial de un problema artificial. En este caso se puede usar el método de la gran M, o bien, el método de las dos fases, para asegurar que el método símplex obtenga una solución óptima para el problema real.

La implantación en computadoras personales del método símplex y sus variantes se han convertido en herramientas tan poderosas que se emplean a menudo para resolver problemas de programación lineal con cientos de miles de restricciones funcionales y variables de decisión y, en ocasiones, para encarar problemas mucho más grandes. Los algoritmos de punto interior proporcionan una nueva herramienta poderosa para resolver problemas muy grandes.

APÉNDICE 4.1 INTRODUCCIÓN AL USO DE LINDO Y LINGO

El paquete LINGO puede aceptar modelos de optimización en cualquiera de los dos estilos o sintaxis: *a*) sintaxis LINDO o *b*) sintaxis LINGO. En primera instancia se descrbirá la sintaxis LINDO. Las ventajas relativas atribuibles a ella consisten en que es muy fácil y natural para enfrentar problemas de programación lineal y con enteros. Su sintaxis se ha utilizado ampliamente desde 1981.

La sintaxis de LINDO permite introducir un modelo en forma natural, en esencia, de la misma forma como se presenta en el libro. Por ejemplo, ésta es una forma sencilla de escribir un modelo en LINDO para el ejemplo de la Wyndor Glass, Co. de la sección 3.1. Bajo el supuesto de que usted ya instaló LINGO, presione sobre el ícono LINGO para arrancar LINGO e inmediatamente después teclee lo siguiente:

```
! Problema de Wyndor Glass Co. Modelo LINDO
! X1 = lotes del producto 1 por semana
! X2 = lotes del producto 2 por semana
! Ganancia, en 1000 dólares

MAX Profit) 3 XI + 5 X2

Subject to

! Tiempo de producción
Plant1) X1 <= 4
Plant2) 2 X2 <= 12
Plant3) 3 X1 + 2 X2 <= 18
END
```

Las primeras cuatro líneas, las cuales comienzan con un signo de admiración, son sólo comentarios. El comentario de la cuarta línea clarifica que la función objetivo está expresada en unidades de miles de dólares. El número 1 000 en este comentario no tiene la coma acostumbrada en frente de los tres últimos dígitos ya que LINDO/LINGO no acepta comas. (La sintaxis de LINDO tampoco acepta paréntesis en expresiones algebraicas.) Las líneas cinco en adelante especifican el modelo. Las variables de decisión pueden especificarse en minúsculas o mayúsculas. Por lo general, las mayúsculas se utilizan con el fin de que las variables no se vean pequeñas con respecto a los "subíndices" que le preceden. En lugar de usar X1 o X2, usted puede utilizar nombres más sugerentes como, por ejemplo, el nombre del producto que se está fabricando; por ejemplo, PUERTAS y VENTANAS, para representar a la variable de decisión a lo largo de todo el modelo.

La quinta línea de la formulación LINDO indica que el objetivo del modelo es maximizar la función objetivo, $3x_1 + 5x_2$. La palabra Profit (ganancia) seguida de un paréntesis es opcional y aclara que la cantidad que se maximiza se llamará Profit en el reporte solución.

El comentario del séptimo renglón señala que las siguientes restricciones son de los tiempos de producción que se usan. Los siguientes tres comienzan por un nombre seguido de un paréntesis para cada restricción funcional. Estas restricciones se escriben de la forma usual excepto por los signos de desigualdad. Como muchos teclados no incluyen los símbolos $\leq$ y $\geq$, LINDO interpreta los símbolos $<$ o $<=$ como $\leq$ y $>$ o $>=$ como $\geq$. (Para teclados que incluyen los símbolos $\leq$ y $\geq$, LINDO no los reconocerá.)

El final de las restricciones se indica con la palabra END. LINDO supone de modo automático las restricciones de no negatividad. Si, por ejemplo, x_1 no tiene restricción de no negatividad, debe indicarse con FREE X1 en el renglón que sigue a END.

Para resolver este modelo en LINGO/LINDO, presione la tecla Bull's Eye rojo en la parte superior de la ventana LINGO. La figura A4.1 muestra el "informe de solución" resultante. Las líneas en la parte superior indican que se encontró una solución óptima o "global" con un valor de la función objetivo de 36 en dos iteraciones. En seguida aparecen los valores de x_1 y x_2 para la solución óptima.

La columna de la derecha de la columna Valores proporciona los **costos reducidos**, los cuales no se han estudiado en este capítulo porque la información que proporcionan se puede derivar del *intervalo permisible* para los coeficientes de la función objetivo. Dichos rangos permisibles se encuentran disponibles (como se verá en la siguiente figura). Cuando la variable es una *variable básica* de la solución óptima (como ambas variables del problema de la Wyndor), su costo reducido es 0 en forma automática. Cuando es una *variable no básica* su costo reducido proporciona información interesante. Una variable cuyo coeficiente objetivo es "muy pequeño" en un modelo de maximización o "muy grande" en un modelo de minimización tendrá un valor de 0 en una solución óptima. El costo reducido indica cuánto se puede *aumentar* (cuando se maximiza) o *disminuir* (cuando se minimiza) este coeficiente antes de que cambie la solución óptima y esta variable se convierta en básica. Sin embargo, recuerde que esta información se obtiene mediante el intervalo permisible para seguir óptimo para el coeficiente de esta variable en la función objetivo. El costo reducido (de una variable no básica) es sólo el *aumento permitido* (cuando se maximiza), a partir del valor actual del coeficiente para quedar dentro de su intervalo permisible o la *reducción permitida* (cuando se minimiza).

La parte inferior de la figura A4.1 proporciona información acerca de las tres restricciones funcionales. La columna Slack o Surplus proporciona la diferencia entre los dos lados de cada restricción. La columna Dual Prices proporciona, con otro nombre, los *precios sombra* de estas restricciones estudiados en la sección

FIGURA A4.1
Informe de solución proporcionado por la sintaxis de LINDO sobre el problema de la Wyndor Glass Co.

```
Global optimal solution found.
Objective value:                                    36.00000
Total solver iterations:
 Variable        Value          Reduced Cost
    X1         2.000000           0.000000
    X2         6.000000           0.000000
   Row     Slack Or Surplus     Dual Prices
  PROFIT       36.00000           1.000000
  PLANT1        2.000000          0.000000
  PLANT2        0.000000          1.500000
  PLANT3        0.000000          1.000000
```

4.7. (Este nombre se debe al hecho, tema que se introducirá en la sección 6.1, de que los precios sombra son los valores óptimos de las variables duales que se presentan en el capítulo 6.) Sin embargo, es importante notar que LINDO utiliza una convención de signos diferente con respecto a la convencional adoptada en cualquier parte de este libro (vea la nota al pie 19 respecto a la definición de precio sombra en la sección 4.7). En particular, en el caso de problemas de minimización, los precios sombra (precios duales) de LINGO/LINDO son los negativos de los nuestros.

Una vez que LINDO le proporcione el reporte de solución, usted tiene la opción de realizar el análisis de rango (sensibilidad). La figura A4.2 muestra el reporte de rango, el cual se genera tecleando en LINGO | Range.

Excepto por el uso de unidades de miles de dólares en lugar de dólares en los coeficientes de la función objetivo, este reporte es idéntico a las últimas tres columnas de las tablas del informe de sensibilidad generado por Excel Solver, que se muestran en la figura 4.10. Así, como se vio en la sección 4.7, los primeros dos renglones de este informe indican que el rango permisible de cada coeficiente de la función objetivo (se supone que no hay otro cambio en el modelo) es

$$0 \leq c_1 \leq 7.5$$
$$2 \leq c_2$$

De manera similar, los últimos tres renglones indican que el intervalo permisible factible en cada lado derecho (en el supuesto de que no hay otro cambio en el modelo) es

$$2 \leq b_1$$
$$6 \leq b_2 \leq 18$$
$$12 \leq b_3 \leq 24$$

Usted puede imprimir los resultados en la forma estándar de Windows tecleando en Files | Print.

Éstos son los fundamentos para comenzar a usar LINGO/LINDO. Usted puede encender o apagar la generación de reportes. Por ejemplo, si se apaga la generación automática del reporte de solución estándar (modo Terse), usted lo puede volver a encender tecleando: LINGO | Options | Interface | Output level | Verbose | Apply. La facilidad de generar reportes de intervalo puede encenderse o apagarse tecleando: LINGO | Options | General solver | Dual computations | Prices & Ranges | Apply.

El segundo estilo de entrada con el que LINGO cuenta es sintaxis LINGO, la cual es mucho más poderosa que la sintaxis LINDO. Las ventajas del uso de la sintaxis LINGO son: *a*) permite expresiones matemáticas arbitrarias incluyendo paréntesis así como todas las operaciones matemáticas comunes tales como la división, la multiplicación, logaritmos, seno, etc., *b*) la facilidad de no sólo resolver problemas de programación lineal sino también no lineal, *c*) la escalabilidad de aplicaciones extensas usando variables con subíndices y conjuntos, *d*) la facilidad de leer datos de entrada a partir de una hoja de cálculo o base de datos y enviar información acerca de la solución a la hoja de cálculo o base de datos, *e*) la facilidad de representar de forma natural relaciones dispersas, *f*) la facilidad de programar de tal forma que usted pueda resolver una serie de modelos en forma automática de la misma manera que cuando realiza un análisis paramétrico. Una formulación del problema Wyndor en LINGO utilizando la facilidad subíndices/conjuntos es:

FIGURA A4.2
Informe de intervalos proporcionado por LINDO sobre el problema de la Wyndor Glass Co.

```
Ranges in which the basis is unchanged:
               Objective   Coefficient     Ranges
                Current     Allowable     Allowable
  Variable    Coefficient    Increase      Decrease
     X1         3.000000     4.500000      3.000000
     X2         5.000000     INFINITY      3.000000
              Righthand    Side Ranges
    Row         Current     Allowable     Allowable
                  RHS        Increase      Decrease
   PLANT1       4.000000     INFINITY      2.000000
   PLANT2      12.000000     6.000000      6.000000
   PLANT3      18.000000     6.000000      6.000000
```

```
! Wyndor Glass Co. Problem;
SETS:
 PRODUCT: PPB, X;           ! Each product has a profit/batch
and amount;
 RESOURCE: HOURSAVAILABLE;   ! Each resource has a capacity;
! Each resource product combination has an hours/batch;
 RXP(RESOURCE,PRODUCT): HPB;
ENDSETS
DATA:
 PRODUCT  = DOORS  WINDOWS;     ! The products;
     PPB =   3       5;         ! Profit per batch;

      RESOURCE = PLANT1 PLANT2 PLANT3;
 HOURSAVAILABLE =   4      12     18;

  HPB =   1  0   ! Hours per batch;
          0  2
          3  2;
ENDDATA
 ! Sum over all products j the profit per batch times batches
produced;
 MAX = @SUM( PRODUCT(j): PPB(j)*X(j));

 @FOR( RESOURCE(i)):  ! For each resource i...;
   ! Sum over all products j of hours per batch time batches
produced...;
   @SUM(RXP(i,j): HPB(i,j)*X(j)) <= HOURSAVAILABLE(i);
     );
```

El problema Wyndor original tiene dos productos y tres recursos. Si Wyndor se expande para tener cuatro productos y cinco recursos, representa un cambio trivial insertar los nuevos datos en la sección DATA. La formulación del modelo se ajusta de manera automática. La facilidad subscript/sets también permite representar modelos en tres dimensiones o en más de manera natural. El problema descrito en la sección 3.6 tiene cinco dimensiones: plantas, máquinas, productos, regiones/clientes y periodos. Lo anterior resulta difícil de representar en una hoja de datos bidimensional, pero es sencillo hacerlo en un lenguaje de modelación con conjuntos y subíndices. En la práctica, para el caso de problemas como el de la sección 3.6, gran parte de las 10(10)(10)(10)(10) = 100 000 combinaciones posibles de relaciones no existen; por ejemplo, no todas las plantas pueden fabricar todos los productos y no todos los clientes demandan todos los productos. La facilidad subscript/sets de los lenguajes de modelación simplifican la representación de dichas relaciones dispersas.

En la mayoría de los modelos que usted ingrese, LINGO podrá detectar de manera automática el uso de sintaxis LINDO o LINGO. Usted puede seleccionar su sintaxis por omisión tecleando: LINGO | Options | Interface | File format | lng (para el caso de LINGO) o ltx (para el caso de LINDO).

LINGO incluye un menú de ayuda en línea con el fin de brindar más detalles y ejemplos. Los suplementos 1 y 2 del capítulo 3 (que se muestran en el sitio de internet del libro) proporcionan una introducción casi completa de LINGO. El tutorial de LINGO en el sitio de internet también proporciona detalles adicionales. Los archivos LINGO/LINDO que se encuentran en el sitio de internet de diferentes capítulos muestran formulaciones LINDO/LINGO de numerosos ejemplos pertenecientes a una gran parte de los capítulos.

REFERENCIAS SELECCIONADAS

1. Bixby, R. E., "Solving Real-World Linear Programs: A Decade and More of Progress", en *Operations Research*, **50**(1): 3-15, enero-febrero de 2002.
2. Dantzig, G. B. y M. N. Thapa, *Linear Programming 1: Introduction*, Springer, Nueva York, 1997.
3. Fourer, R., "Software Survey: Linear Programming", en *OR/MS Today*, junio de 2007, pp. 42-51.
4. Luenberger, D. y Y. Ye, *Linear and Nolinear Programming*, 3a. ed., Springer, Nueva York, 2008.

5. Maros, I., *Computational Techniques of the Simplex Method*, Kluwer Academic Publishers (en la actualidad, Springer), Boston, MA, 2003.
6. Schrage, L., *Optimization Modeling with LINGO*, LINDO Systems, Chicago, 2008.
7. Tretkoff, C. e I. Lustig, "New Age of Optimization Applications", *OR/MS Today*, diciembre de 2006, pp. 46-49.
8. Venderbei, R. J., *Linear Programming: Foundations and Extensions*, 3a. ed., Springer, Nueva York, 2008.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN NUESTRO SITIO DE INTERNET (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos para el capítulo 4

Ejemplos de demostración en OR Tutor:

Interpretación de variables de holgura (Interpretation of Slack Variables)
Forma algebraica del método símplex (Simplex Method—Algebraic Form)
Forma tabular del método símplex (Simplex Method—Tabular Form)

Rutinas interactivas en IOR Tutorial:

Introducción o revisión de un problema general de programación lineal (Enter or Revise a General Linear Programming Model)
Preparación para el método símplex, sólo interactivo (Set Up for the Simplex Method—Interactive Only)
Solución interactiva por el método símplex (Solve Interactively by the Simplex Method)
Método gráfico interactivo (Interactive Graphical Method)

Rutinas automáticas en IOR Tutorial:

Solución automática por el método símplex (Solve Automatically by the Simplex Method)
Solución automática por el algoritmo de punto interior (Solve Automatically by the Interior-Point Algorithm)
Método gráfico y análisis de sensibilidad (Graphical Method and Sensitivity Analysis)

Complemento de Excel:

Premium Solver for Education

Archivos (capítulo 3) para resolver los ejemplos de la Wyndor y de terapia de radiación:

Archivos de Excel
Archivo de LINGO/LINDO
Archivo de MPL/CPLEX

Glosario del capítulo 4

Vea en el apéndice 1 la documentación del software.

PROBLEMAS

Los símbolos de la izquierda de algunos problemas (o de sus incisos) significan lo siguiente:

D: El ejemplo de demostración correspondiente de la lista de la página precedente puede ser útil.
I: Se sugiere que use la rutina interactiva correspondiente enumerada en la página anterior (la impresión registra su trabajo).
C: Use la computadora con cualquiera de las opciones de software disponibles (o según lo indique su instructor) para resolver el problema de manera automática. (Vea en la sección 4.8 una lista de las opciones que contiene este libro y en el sitio en internet del libro.)

Un asterisco en el número del problema indica que al final del libro se da al menos una respuesta parcial.

4.1-1. Considere el siguiente problema.

$$\text{Maximizar} \quad Z = x_1 + 2x_2,$$

sujeta a

$$\begin{aligned} x_1 \quad &\leq 5 \\ x_2 &\leq 6 \\ x_1 + x_2 &\leq 8 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

a) Grafique la región factible y marque con un círculo las soluciones FEV.
b) En cada solución FEV identifique el par de ecuaciones de fronteras de restricción que satisface.
c) En cada solución FEV utilice este par de ecuaciones de fronteras de restricción para obtener la solución algebraica de los valores de x_1 y x_2 en el vértice.
d) En cada solución FEV, identifique sus soluciones FEV adyacentes.
e) En cada par de soluciones FEV adyacentes, identifique con su ecuación la frontera de restricción común.

4.1-2. Considere el siguiente problema.

$$\text{Maximizar} \quad Z = 3x_1 + 2x_2,$$

sujeta a

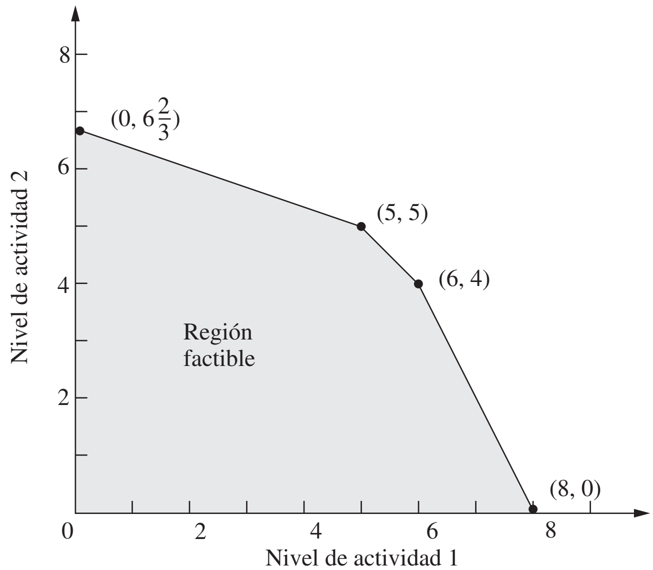
$$\begin{aligned} 2x_1 + x_2 &\leq 6 \\ x_1 + 2x_2 &\leq 6 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

D,I ***a***) Use el método gráfico para resolver este problema. Marque con un círculo los vértices de la gráfica.
b) En cada solución FEV, identifique el par de ecuaciones de fronteras de restricción que satisface.
c) En cada solución FEV, identifique sus soluciones FEV adyacentes.
d) Calcule Z en cada solución FEV. Use esta información para identificar la solución óptima.
e) Describa en la gráfica lo que hace el método símplex paso a paso para resolver este problema.

4.1-3. Cierto modelo de programación lineal con dos actividades tiene la región factible que se muestra.



El objetivo es maximizar la ganancia total de las dos actividades. La ganancia unitaria de la actividad 1 asciende a \$1 000 y la de la actividad 2 a \$2 000.

a) Calcule la ganancia total según cada solución FEV. Use esta información para encontrar la solución óptima.
b) Utilice los conceptos de solución del método símplex que se presentaron en la sección 4.1 para identificar la secuencia de soluciones FEV que examinará el método símplex para llegar a la solución óptima.

4.1-4.* Considere el modelo de programación lineal (que se explicó al final del libro) que se formuló para el problema 3.2-3.

a) Utilice el análisis gráfico para identificar todas las *soluciones en los vértices* en este modelo. Diga si cada una de ellas es factible o no factible.
b) Calcule el valor de la función objetivo de cada una de las soluciones FEV. Utilice esta información para identificar una solución óptima.
c) Utilice los conceptos de solución del método símplex que se presentaron en la sección 4.1 para identificar la secuencia de soluciones FEV que examinará el método símplex para llegar a una solución óptima. (*Sugerencia:* Existen dos secuencias alternativas que se pueden identificar en este modelo en particular.)

4.1-5. Repita el problema 4.1-4 en el siguiente problema.

$$\text{Maximizar} \quad Z = x_1 + 2x_2,$$

sujeta a

$$\begin{aligned} x_1 + 3x_2 &\leq 8 \\ x_1 + x_2 &\leq 4 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

4.1-6. Describa en una gráfica lo que hace el método símplex paso a paso para resolver el siguiente problema.

Maximizar $Z = 2x_1 + 3x_2$,

sujeta a

$$\begin{aligned} -3x_1 + x_2 &\leq 1 \\ 4x_1 + 2x_2 &\leq 20 \\ 4x_1 - x_2 &\leq 10 \\ -x_1 + 2x_2 &\leq 5 \end{aligned}$$

y

$x_1 \geq 0, \quad x_2 \geq 0.$

4.1-7. Describa en una gráfica lo que hace el método símplex paso a paso para resolver el siguiente problema.

Minimizar $Z = 5x_1 + 7x_2$,

sujeta a

$$\begin{aligned} 2x_1 + 3x_2 &\geq 147 \\ 3x_1 + 4x_2 &\geq 210 \\ x_1 + x_2 &\geq 63 \end{aligned}$$

y

$x_1 \geq 0, \quad x_2 \geq 0.$

4.1-8. Diga si cada una de las siguientes afirmaciones sobre los problemas de programación lineal es falsa o verdadera y justifique su respuesta.

a) En problemas de minimización, si se evalúa la función objetivo en una solución FEV y no es mayor que su valor en las soluciones FEV adyacentes, entonces esa solución es óptima.

b) Sólo las soluciones FEV pueden ser óptimas, de manera que el número de soluciones óptimas no puede ser mayor que el número de soluciones FEV.

c) Si existen soluciones óptimas múltiples, entonces una solución FEV óptima puede tener una solución FEV adyacente que también es óptima (que da como resultado el mismo valor de Z).

4.1-9. Las siguientes afirmaciones son paráfrasis inexactas de los seis conceptos de solución de la sección 4.1. En cada caso explique qué está mal.

a) La mejor solución FEV siempre es óptima.

b) Una iteración del método símplex verifica si la solución FEV actual es óptima y, si no lo es, se desplaza hacia una nueva solución FEV.

c) Aunque se puede elegir cualquier solución FEV como solución FEV inicial, el método símplex siempre comienza en el origen.

d) Cuando el método símplex está listo para elegir una nueva solución FEV para moverse de la solución FEV actual, sólo considera las soluciones adyacentes porque es probable que una de ellas sea óptima.

e) Para elegir la nueva solución FEV y moverse de la solución FEV actual, el método símplex identifica todas las soluciones FEV adyacentes y determina cuál proporciona la mayor tasa de mejoramiento del valor de la función objetivo.

4.2-1. Reconsidere el modelo del problema 4.1-4.

a) Introduzca las variables de holgura a fin de escribir las restricciones funcionales en la forma aumentada.

b) En cada solución FEV, identifique la solución BF correspondiente al calcular los valores de las variables de holgura. En cada solución BF, use los valores de las variables para identificar las variables no básicas y las básicas.

c) En cada solución BF demuestre (mediante la sustitución de la solución) que después de hacer las variables no básicas iguales a cero, esta solución BF también es la solución simultánea del sistema de ecuaciones que obtuvo en el inciso *a*).

4.2-2. Reconsidere el modelo del problema 4.1-5. Siga las instrucciones del problema 4.2-1 para los incisos *a*), *b*) y *c*).

d) Repita el inciso *b*) para las soluciones no factibles en los vértices y las soluciones básicas no factibles correspondientes.

e) Repita el inciso *c*) para las soluciones básicas no factibles.

4.3-1. Lea la referencia que describe todo el estudio de investigación de operaciones que se resume en el Recuadro de aplicación que se muestra en la sección 4.3. Describa en forma breve la aplicación del método símplex en este estudio. Después, haga una lista con los diferentes beneficios financieros y no financieros que resulten de este estudio.

D,I **4.3-2.** Aplique el método símplex (en forma algebraica) paso a paso para resolver el modelo del problema 4.1-4.

4.3-3. Reconsidere el modelo del problema 4.1-5.

a) Aplique el método símplex (en su forma algebraica) a mano para resolver este modelo.

D,I ***b***) Repita el inciso *a*) con la rutina interactiva correspondiente en el IOR Tutorial.

C ***c***) Verifique la solución óptima que obtuvo por medio de un paquete de software basado en el método símplex.

D,I **4.3-4.*** Utilice el método símplex (en su forma algebraica) paso a paso para resolver el siguiente problema.

Maximizar $Z = 4x_1 + 3x_2 + 6x_3$,

sujeta a

$$\begin{aligned} 3x_1 + x_2 + 3x_3 &\leq 30 \\ 2x_1 + 2x_2 + 3x_3 &\leq 40 \end{aligned}$$

y

$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$

D,I **4.3-5.** Use el método símplex (en forma algebraica) paso a paso para resolver el siguiente problema.

Maximizar $Z = 3x_1 + 4x_2 + 5x_3$,

sujeta a

$$\begin{aligned} 3x_1 + x_2 + 5x_3 &\leq 150 \\ x_1 + 4x_2 + x_3 &\leq 120 \\ 2x_1 + 2x_3 &\leq 105 \end{aligned}$$

y

$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$

4.3-6. Considere el siguiente problema.

Maximizar $Z = 5x_1 + 3x_2 + 4x_3$,

sujeta a

$$\begin{aligned} 2x_1 + x_2 + x_3 &\leq 20 \\ 3x_1 + x_2 + 2x_3 &\leq 30 \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

Usted obtiene la información de que las variables *diferentes de cero* en la solución óptima son x_2 y x_3.

a) Describa cómo se puede usar la información para adaptar el método símplex a fin de resolver este problema en el menor número posible de iteraciones (si comienza con la solución BF inicial de costumbre). *No* realice iteraciones.

b) Utilice el procedimiento desarrollado en el inciso *a*) para resolver este problema a mano. (*No* use el OR Courseware.)

4.3-7. Considere el siguiente problema.

$$\text{Maximizar} \quad Z = 2x_1 + 4x_2 + 3x_3,$$

sujeta a

$$\begin{aligned} x_1 + 3x_2 + 2x_3 &\leq 30 \\ x_1 + x_2 + x_3 &\leq 24 \\ 3x_1 + 5x_2 + 3x_3 &\leq 60 \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

Usted obtiene la información de que $x_1 > 0$, $x_2 = 0$ y $x_3 > 0$ en la solución óptima.

a) Describa cómo puede usar la información para adaptar el método símplex a fin de resolver este problema en el número mínimo posible de iteraciones (si comienza con la solución BF inicial de costumbre). *No* realice iteraciones.

b) Utilice el procedimiento desarrollado en el inciso *a*) para resolver este problema a mano. (*No* use el OR Courseware.)

4.3-8. Diga si cada una de las siguientes afirmaciones es falsa o verdadera y justifique su respuesta mediante referencia a las citas específicas del capítulo.

a) La regla del método símplex para elegir la variable básica entrante se usa porque siempre conduce a la mejor solución BF adyacente (mayor valor de Z).

b) La regla del cociente mínimo del método símplex para elegir la variable básica que sale se usa porque si se hace otra elección con un cociente mayor se llega a una solución que no es factible.

c) Cuando el método símplex obtiene la siguiente solución BF se usan operaciones algebraicas elementales para eliminar cada variable no básica de todas las ecuaciones menos una (*su* ecuación) y para darle un coeficiente de +1 en esa ecuación.

D,I **4.4-1.** Repita el problema 4.3-2; utilice la forma tabular del método símplex.

D,I,C **4.4-2.** Repita el problema 4.3-3; utilice la forma tabular del método símplex.

4.4-3. Considere el siguiente problema.

$$\text{Maximizar} \quad Z = 2x_1 + x_2,$$

sujeta a

$$\begin{aligned} x_1 + x_2 &\leq 40 \\ 4x_1 + x_2 &\leq 100 \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

a) Resuelva este problema de manera gráfica y a mano. Identifique todas las soluciones FEV.

D,I ***b***) Ahora use el IOR Tutorial para resolver este problema en forma gráfica.

D ***c***) Mediante cálculos manuales resuelva este problema por el método símplex en forma algebraica.

D,I ***d***) Ahora use el IOR Tutorial para resolver este problema de manera interactiva por el método símplex en su forma algebraica.

D ***e***) Mediante cálculos manuales resuelva este problema por el método símplex en forma tabular.

D,I ***f***) Ahora utilice el IOR Tutorial para resolver este problema de manera interactiva por el método símplex en forma tabular.

C ***g***) Use un paquete de software basado en el método símplex para resolver este problema.

4.4-4. Repita el problema 4.4-3 para resolver el siguiente problema.

$$\text{Maximizar} \quad Z = 2x_1 + 3x_2,$$

sujeta a

$$\begin{aligned} x_1 + 2x_2 &\leq 30 \\ x_1 + x_2 &\leq 20 \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

4.4-5. Considere el siguiente problema.

$$\text{Maximizar} \quad Z = 5x_1 + 9x_2 + 7x_3,$$

sujeta a

$$\begin{aligned} x_1 + 3x_2 + 2x_3 &\leq 10 \\ 3x_1 + 4x_2 + 2x_3 &\leq 12 \\ 2x_1 + x_2 + 2x_3 &\leq 8 \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

D,I ***a***) Aplique el método símplex paso a paso en la forma algebraica.

D,I ***b***) Aplique el método símplex paso a paso en la forma tabular.

C ***c***) Use un paquete de software basado en el método símplex para resolver el problema.

4.4-6. Considere el siguiente problema.

$$\text{Maximizar} \quad Z = 3x_1 + 5x_2 + 6x_3,$$

sujeta a

$$\begin{aligned} 2x_1 + x_2 + x_3 &\leq 4 \\ x_1 + 2x_2 + x_3 &\leq 4 \\ x_1 + x_2 + 2x_3 &\leq 4 \\ x_1 + x_2 + x_3 &\leq 3 \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

D,I ***a***) Aplique el método símplex paso a paso en forma algebraica.

D,I ***b***) Aplique el método símplex en forma tabular.

C ***c***) Use un paquete de computadora basado en el método símplex para resolver el problema.

D,I **4.4-7.** Aplique el método símplex paso a paso (en forma tabular) para resolver el siguiente problema.

Maximizar $Z = 2x_1 - x_2 + x_3$,

sujeta a

$$3x_1 + x_2 + x_3 \le 6$$
$$x_1 - x_2 + 2x_3 \le 1$$
$$x_1 + x_2 - x_3 \le 2$$

y

$$x_1 \ge 0, \quad x_2 \ge 0, \quad x_3 \ge 0.$$

D,I **4.4-8.** Aplique el método símplex paso a paso para resolver el siguiente problema.

Maximizar $Z = -x_1 + x_2 + 2x_3$,

sujeta a

$$x_1 + 2x_2 - x_3 \le 20$$
$$-2x_1 + 4x_2 + 2x_3 \le 60$$
$$2x_1 + 3x_2 + x_3 \le 50$$

y

$$x_1 \ge 0, \quad x_2 \ge 0, \quad x_3 \ge 0.$$

4.5-1. Considere las siguientes afirmaciones sobre programación lineal y el método símplex. Diga si cada una es falsa o verdadera y justifique su respuesta.

a) En una iteración específica del método símplex, si hay un empate de la variable que debe salir de la base, entonces la siguiente solución BF debe tener al menos una variable básica igual a cero.
b) Si no existe una variable básica saliente en alguna iteración, entonces el problema no tiene soluciones factibles.
c) Si al menos una de las variables básicas tiene un coeficiente igual a cero en el renglón 0 de la tabla símplex final, entonces el problema tiene soluciones óptimas múltiples.
d) Si el problema tiene soluciones óptimas múltiples, entonces el problema debe tener una región factible acotada.

4.5-2. Suponga que se proporcionan las siguientes restricciones para un modelo de programación lineal con variables de decisión x_1 y x_2.

$$-2x_1 + 3x_2 \le 12$$
$$-3x_1 + 2x_2 \le 12$$

y

$$x_1 \ge 0, \quad x_2 \ge 0.$$

a) Demuestre en una gráfica que la región factible es no acotada.
b) Si el objetivo es maximizar $Z = -x_1 + x_2$, ¿tiene el modelo una solución óptima? Si es así, encuéntrela. Si no, explique por qué.
c) Repita el inciso *b*) si el objetivo es maximizar $Z = x_1 - x_2$.
d) En ciertas funciones objetivo, este modelo no tiene solución óptima. ¿Significa ello que no hay soluciones buenas según el modelo? Explique. ¿Qué pudo haber salido mal al formular el modelo?
D,I ***e***) Seleccione una función objetivo para la que este modelo no tenga solución óptima. Después, aplique el método símplex paso a paso para demostrar que Z es no acotada.
C ***f***) En el caso de la función objetivo seleccionada en el inciso *e*), aplique un paquete de software basado en el método símplex para determinar que Z no es acotada.

4.5-3. Siga las instrucciones del problema 4.5-2 con las siguientes restricciones:

$$2x_1 - x_2 \le 20$$
$$x_1 - 2x_2 \le 20$$

y

$$x_1 \ge 0, \quad x_2 \ge 0.$$

D,I **4.5-4.** Considere el siguiente problema.

Maximizar $Z = 5x_1 + x_2 + 3x_3 + 4x_4$,

sujeta a

$$x_1 - 2x_2 + 4x_3 + 3x_4 \le 20$$
$$-4x_1 + 6x_2 + 5x_3 - 4x_4 \le 40$$
$$2x_1 - 3x_2 + 3x_3 + 8x_4 \le 50$$

y

$$x_1 \ge 0, \quad x_2 \ge 0, \quad x_3 \ge 0, \quad x_4 \ge 0.$$

Utilice el método símplex para demostrar que Z no está acotada.

4.5-5. Una propiedad básica de cualquier problema de programación lineal con una región factible acotada es que toda solución factible se puede expresar como una combinación convexa de las soluciones FEV (tal vez en más de una forma). De igual manera, en la forma aumentada del problema, toda solución factible se puede expresar como una combinación convexa de las soluciones BF.

a) Demuestre que *cualquier* combinación convexa de *cualquier* conjunto de soluciones factibles debe ser una solución factible (de manera que cualquier combinación convexa de soluciones FEV debe ser factible).
b) Utilice el resultado establecido en el inciso *a*) para demostrar que cualquier combinación convexa de soluciones BF debe ser una solución factible.

4.5-6. Utilice los datos del problema 4.5-5 para demostrar que las siguientes afirmaciones deben ser ciertas para cualquier problema de programación lineal que tiene una región factible acotada y soluciones óptimas múltiples:

a) Toda combinación convexa de soluciones BF óptimas debe ser óptima.
b) Ninguna otra solución factible puede ser óptima.

4.5-7. Considere un problema de programación lineal de dos variables cuyas soluciones FEV son (0, 0), (6, 0), (6, 3), (3, 3) y (0, 2). (Vea en el problema 3.2-2 la gráfica de la región factible.)

a) Use la gráfica de la región factible para identificar todas las restricciones del modelo.
b) Para cada par de soluciones FEV adyacentes, dé un ejemplo de una función objetivo tal que todos los puntos del segmento entre esas dos soluciones en los vértices sean soluciones óptimas.
c) Ahora suponga que la función objetivo es $Z = -x_1 + 2x_2$. Aplique el método gráfico para encontrar todas las soluciones óptimas.
D,I ***d***) En el caso de la función objetivo del inciso *c*), aplique el método símplex paso a paso para encontrar todas las soluciones BF óptimas. Después escriba una expresión algebraica que identifique todas las soluciones óptimas.

D,I **4.5-8.** Considere el siguiente problema.

Maximizar $Z = 50x_1 + 25x_2 + 20x_3 + 40x_4,$

sujeta a

$$2x_1 + x_2 \leq 30$$
$$x_3 + 2x_4 \leq 20$$

y

$x_j \geq 0,$ para $j = 1, 2, 3, 4.$

Trabaje el método símplex paso a paso para encontrar todas las soluciones BF óptimas.

4.6-1.* Considere el siguiente problema.

Maximizar $Z = 2x_1 + 3x_2,$

sujeta a

$$x_1 + 2x_2 \leq 4$$
$$x_1 + x_2 = 3$$

y

$x_1 \geq 0,$ $x_2 \geq 0.$

D,I ***a***) Resuelva este problema en forma gráfica.
b) Aplique el método de la gran *M* para construir la primera tabla símplex completa para el método símplex e identificar la solución BF inicial (artificial) correspondiente. También identifique la variable básica entrante y la variable básica saliente.
I ***c***) A partir del inciso *b*) trabaje el método símplex paso a paso para resolver el problema.

4.6-2. Considere el siguiente problema.

Maximizar $Z = 4x_1 + 2x_2 + 3x_3 + 5x_4,$

sujeta a

$$2x_1 + 3x_2 + 4x_3 + 2x_4 = 300$$
$$8x_1 + x_2 + x_3 + 5x_4 = 300$$

y

$x_j \geq 0,$ para $j = 1, 2, 3, 4.$

a) Aplique el método de la gran *M* para construir la primera tabla símplex completa para el método símplex e identifique la solución BF inicial (artificial) correspondiente. Identifique también la variable básica entrante inicial y la variable básica saliente.
I ***b***) Aplique el método símplex paso a paso para resolver el problema.
c) Utilice el método de las dos fases para construir la primera tabla símplex completa para la fase 1 e identifique la solución BF inicial (artificial) correspondiente. Identifique también la variable básica entrante inicial y la variable básica saliente.
I ***d***) Aplique la fase 1 paso a paso.
e) Construya la primera tabla símplex completa de la fase 2.
I ***f***) Aplique la fase 2 paso a paso para resolver el problema.
g) Compare la serie de soluciones BF que obtuvo en el inciso *b*) con las de los incisos *d*) y *f*). ¿Cuáles de estas soluciones son factibles sólo para el problema artificial que se obtuvo al introducir las variables artificiales y cuáles son de hecho factibles para el problema real?
C ***h***) Utilice un paquete de software basado en el método símplex para resolver el problema.

4.6-3.* Considere el siguiente problema.

Minimizar $Z = 2x_1 + 3x_2 + x_3,$

sujeta a

$$x_1 + 4x_2 + 2x_3 \geq 8$$
$$3x_1 + 2x_2 \geq 6$$

y

$x_1 \geq 0,$ $x_2 \geq 0,$ $x_3 \geq 0.$

a) Reformule este problema para que se ajuste a nuestra forma estándar del modelo de programación lineal que se presentó en la sección 3.2.
I ***b***) Utilice el método de la gran *M* para aplicar el método símplex paso a paso y resolver el problema.
I ***c***) Utilice el método de las dos fases para aplicar el método símplex paso a paso y resolver el problema.
d) Compare la secuencia de soluciones BF que obtuvo en los incisos *b*) y *c*). ¿Cuáles de estas soluciones son factibles sólo para el problema artificial que obtuvo al introducir las variables artificiales y cuáles son de hecho factibles para el problema real?
C ***e***) Utilice un software basado en el método símplex para resolver el problema.

4.6-4. De acuerdo con el método de la gran *M*, explique por qué el método símplex nunca elegiría una variable artificial como variable básica entrante, una vez que todas las variables artificiales son no básicas.

4.6-5. Considere el siguiente problema.

Maximizar $Z = 5x_1 + 4x_2,$

sujeta a

$$3x_1 + 2x_2 \leq 6$$
$$2x_1 - x_2 \geq 6$$

y

$x_1 \geq 0,$ $x_2 \geq 0.$

a) Demuestre en una gráfica que este problema no tiene soluciones factibles.
C ***b***) Utilice un paquete de computadora basado en el método símplex para determinar que el problema no tiene soluciones factibles.
I ***c***) Utilice el método de la gran *M* para aplicar el método símplex paso a paso y demostrar que el problema no tiene soluciones factibles.
I ***d***) Repita el inciso *c*) para la fase 1 del método de las dos fases.

4.6-6. Siga las instrucciones del problema 4.6-5 para solucionar el siguiente problema.

Minimizar $Z = 5\,000x_1 + 7\,000x_2,$

sujeta a

$$-2x_1 + x_2 \geq 1$$
$$x_1 - 2x_2 \geq 1$$

y

$x_1 \geq 0,$ $x_2 \geq 0.$

4.6-7. Considere el siguiente problema.

Maximizar $Z = 2x_1 + 5x_2 + 3x_3$,

sujeta a

$$x_1 - 2x_2 + x_3 \geq 20$$
$$2x_1 + 4x_2 + x_3 = 50$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

a) Utilice el método de la gran *M* para construir la primera tabla símplex completa para el método símplex e identifique la solución BF inicial (artificial) correspondiente. También identifique la variable básica entrante inicial y la variable básica saliente.

I ***b***) Aplique el método símplex paso a paso para resolver el problema.

I ***c***) Utilice el método de las dos fases para construir la primera tabla símplex completa para la fase 1 e identifique la solución BF inicial (artificial) correspondiente. También identifique la variable básica entrante inicial y la variable básica saliente.

I ***d***) Aplique la fase 1 paso a paso.

e) Construya la primera tabla símplex completa de la fase 2.

I ***f***) Aplique la fase 2 paso a paso para resolver el problema.

g) Compare la secuencia de soluciones BF que obtuvo en el inciso *b*) con las de los incisos *d*) y *f*). ¿Cuáles de estas soluciones son factibles sólo para el problema artificial que obtuvo al introducir las variables artificiales y cuáles son factibles para el problema real?

C ***h***) Utilice un paquete de software basado en el método símplex para resolver el problema.

4.6-8. Considere el siguiente problema.

Minimizar $Z = 2x_1 + x_2 + 3x_3$,

sujeta a

$$5x_1 + 2x_2 + 7x_3 = 420$$
$$3x_1 + 2x_2 + 5x_3 \geq 280$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

I ***a***) Utilice el método de las dos fases y aplique la fase 1 paso a paso.

C ***b***) Emplee un paquete de software basado en el método símplex para formular y resolver la fase 1 del problema.

I ***c***) Aplique la fase 2 paso a paso para resolver el problema original.

C ***d***) Utilice un programa de computadora basado en el método símplex para resolver el problema original.

4.6-9.* Considere el siguiente problema.

Minimizar $Z = 3x_1 + 2x_2 + 4x_3$,

sujeta a

$$2x_1 + x_2 + 3x_3 = 60$$
$$3x_1 + 3x_2 + 5x_3 \geq 120$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

I ***a***) Utilice el método de la gran *M* para aplicar el método símplex paso a paso a fin de resolver el problema.

I ***b***) Emplee el método de las dos fases para aplicar el método símplex paso a paso y resolver el problema.

c) Compare la serie de soluciones BF de los incisos *a*) y *b*). ¿Cuáles de estas soluciones son factibles sólo para el problema artificial que se obtuvo al introducir las variables artificiales y cuáles son factibles para el problema real?

C ***d***) Utilice un paquete de software basado en el método símplex para resolver el problema.

4.6-10. Siga las instrucciones del problema 4.6-9 para el siguiente problema.

Minimizar $Z = 3x_1 + 2x_2 + 7x_3$,

sujeta a

$$-x_1 + x_2 = 10$$
$$2x_1 - x_2 + x_3 \geq 10$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

4.6-11. Diga si cada una de las siguientes afirmaciones es falsa o verdadera y justifique su respuesta.

a) Cuando un modelo de programación lineal tiene una restricción de igualdad, se introduce una variable artificial a esta restricción con el fin de comenzar el método símplex con una solución básica inicial obvia que sea factible para el modelo original.

b) Cuando se crea un problema artificial al introducir variables artificiales y al usar el método de la gran *M*, si todas las variables artificiales de una solución óptima del problema artificial son iguales a cero, entonces el problema real no tiene soluciones factibles.

c) En la práctica suele aplicarse el método de las dos fases porque, en general, se requieren menos iteraciones para llegar a una solución óptima que en el método de la gran *M*.

4.6-12. Considere el siguiente problema.

Maximizar $Z = 3x_1 + 7x_2 + 5x_3$,

sujeta a

$$3x_1 + x_2 + 2x_3 \leq 9$$
$$-2x_1 + x_2 + 3x_3 \leq 12$$

y

$$x_2 \geq 0, \quad x_3 \geq 0$$

(sin restricción de no negatividad sobre x_1).

a) Reformule este problema para que todas las variables tengan restricciones de no negatividad.

D,I ***b***) Aplique el método símplex paso a paso para resolver el problema.

C ***c***) Use un paquete de software basado en el método símplex para resolver el problema.

4.6-13.* Considere el siguiente problema.

Maximizar $Z = -x_1 + 4x_2$,

sujeta a

$$-3x_1 + x_2 \leq 6$$
$$x_1 + 2x_2 \leq 4$$
$$x_2 \geq -3$$

(sin restricción de cota inferior para x_1).

D,I ***a***) Resuelva este problema por el método gráfico.
b) Reformule este problema de manera que tenga sólo dos restricciones funcionales y todas las variables tengan restricciones de no negatividad.
D,I ***c***) Aplique el método símplex paso a paso para resolver este problema.

4.6-14. Considere el siguiente problema.

$$\text{Maximizar} \quad Z = -x_1 + 2x_2 + x_3,$$

sujeta a

$$\begin{aligned} 3x_2 + x_3 &\le 120 \\ x_1 - x_2 - 4x_3 &\le 80 \\ -3x_1 + x_2 + 2x_3 &\le 100 \end{aligned}$$

(sin restricciones de no negatividad).
a) Reformule este problema para que todas las variables tengan restricciones de no negatividad.
D,I ***b***) Aplique el método símplex paso a paso para resolver este problema.
C ***c***) Utilice un paquete de computadora basado en el método símplex para resolver el problema.

4.6-15. En este capítulo se describió el método símplex según se aplica a problemas de programación lineal en los que la función objetivo se debe maximizar. En la sección 4.6 se explicó cómo convertir un problema de minimización en uno equivalente de maximización para aplicar el método símplex. Otra opción en los problemas de minimización es hacer algunas modificaciones a las instrucciones que se dieron para el método símplex, a fin de aplicar el algoritmo en forma directa.
a) Describa cuáles deberían ser estas modificaciones.
b) Utilice el método de la gran *M* para aplicar el algoritmo modificado que desarrolló en el inciso *a*) y resuelva el siguiente problema de manera directa y a mano. (*No* use el OR Courseware.)

$$\text{Minimizar} \quad Z = 3x_1 + 8x_2 + 5x_3,$$

sujeta a

$$\begin{aligned} 3x_2 + 4x_3 &\ge 70 \\ 3x_1 + 5x_2 + 2x_3 &\ge 70 \end{aligned}$$

y

$$x_1 \ge 0, \quad x_2 \ge 0, \quad x_3 \ge 0.$$

4.6-16. Considere el siguiente problema.

$$\text{Maximizar} \quad Z = -2x_1 + x_2 - 4x_3 + 3x_4,$$

sujeta a

$$\begin{aligned} x_1 + x_2 + 3x_3 + 2x_4 &\le 4 \\ x_1 - x_3 + x_4 &\ge -1 \\ 2x_1 + x_2 &\le 2 \\ x_1 + 2x_2 + x_3 + 2x_4 &= 2 \end{aligned}$$

y

$$x_2 \ge 0, \quad x_3 \ge 0, \quad x_4 \ge 0$$

(sin restricciones de no negatividad para x_1).
a) Formule de nuevo este problema para que se ajuste a nuestra forma estándar del modelo de programación lineal que se presentó en la sección 3.2.
b) Utilice el método de la gran *M* para construir la primera tabla símplex completa para el método símplex e identifique la solución BF inicial (artificial). Identifique también la variable básica entrante y la variable básica saliente.
c) Use el método de las dos fases para construir el renglón 0 de la primera tabla símplex de la fase 1.
C ***d***) Utilice un paquete de computadora basado en el método símplex para resolver este problema.

I **4.6-17.** Considere el siguiente problema.

$$\text{Maximizar} \quad Z = 4x_1 + 5x_2 + 3x_3,$$

sujeta a

$$\begin{aligned} x_1 + x_2 + 2x_3 &\ge 20 \\ 15x_1 + 6x_2 - 5x_3 &\le 50 \\ x_1 + 3x_2 + 5x_3 &\le 30 \end{aligned}$$

y

$$x_1 \ge 0, \quad x_2 \ge 0, \quad x_3 \ge 0.$$

Aplique el método símplex paso a paso para demostrar que este problema no tiene soluciones factibles.

4.7-1. Consulte la figura 4.10 y el *rango permisible* para cada lado derecho de las restricciones del problema de la Wyndor Glass Co. dado en la sección 3.1. Utilice el método gráfico para demostrar que cada intervalo permisible dado es correcto.

4.7-2. Reconsidere el modelo del problema 4.1-5. Interprete los lados derechos de las restricciones funcionales como la cantidad disponible de los recursos respectivos.
I ***a***) Utilice el análisis gráfico para determinar los precios sombra de los recursos respectivos, tal como en la figura 4.8.
I ***b***) Recurra al análisis gráfico para realizar un análisis de sensibilidad de este modelo. En particular, verifique cada parámetro del modelo para determinar si es un parámetro *sensible* (un parámetro cuyo valor no puede cambiar sin que se modifique la solución óptima) mediante un análisis de la gráfica que identifica la solución óptima.
I ***c***) Utilice el análisis gráfico como en la figura 4.9 para determinar el intervalo permisible para cada valor c_j (coeficiente de x_j en la función objetivo) sobre el que la solución óptima actual sigue óptima.
I ***d***) Si se cambia sólo un valor b_i (el lado derecho de la restricción funcional *i*) se mueve la frontera de restricción correspondiente. Si la solución FEV óptima actual está sobre esta frontera de restricción, la solución FEV también se moverá. Aplique el análisis gráfico para determinar el intervalo permisible de cada valor b_i sobre el cual esta solución FEV sigue factible.
C ***e***) Verifique sus respuestas a los incisos *a*), *c*) y *d*) con un paquete de computadora basado en el método símplex para resolver el problema y después generar la información del análisis de sensibilidad.

4.7-3. Se proporciona el siguiente problema de programación lineal.

$$\text{Maximizar} \quad Z = 3x_1 + 2x_2,$$

sujeta a

$$\begin{aligned} 3x_1 &\le 60 \quad &\text{(recurso 1)} \\ 2x_1 + 3x_2 &\le 75 \quad &\text{(recurso 2)} \\ 2x_2 &\le 40 \quad &\text{(resurso 3)} \end{aligned}$$

y

$$x_1 \ge 0, \quad x_2 \ge 0.$$

D,I ***a***) Resuelva este problema en forma gráfica.
b) Mediante el análisis gráfico, encuentre los precios sombra de los recursos.
c) Determine cuántas unidades adicionales del recurso 1 se necesitan para aumentar en 15 el valor óptimo de Z.

4.7-4. Considere el siguiente problema.

$$\text{Maximizar} \quad Z = x_1 - 7x_2 + 3x_3,$$

sujeta a

$$\begin{aligned} 2x_1 + x_2 - x_3 &\leq 4 \quad &\text{(recurso 1)} \\ 4x_1 - 3x_2 &\leq 2 \quad &\text{(recurso 2)} \\ -3x_1 + 2x_2 + x_3 &\leq 3 \quad &\text{(recurso 3)} \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

D,I ***a***) Aplique el método símplex paso a paso para resolver el problema.
b) Identifique los precios sombra de los tres recursos y describa su significado.
C ***c***) Recurra a un software basado en el método símplex para resolver el problema y genere el informe de sensibilidad. Use esta información para determinar los precios sombra de cada recurso, el intervalo permisible para seguir óptimo de cada coeficiente de la función objetivo y el intervalo permisible para seguir factible de cada lado derecho.

4.7-5.* Considere el siguiente problema.

$$\text{Maximizar} \quad Z = 2x_1 - 2x_2 + 3x_3,$$

sujeta a

$$\begin{aligned} -x_1 + x_2 + x_3 &\leq 4 \quad &\text{(recurso 1)} \\ 2x_1 - x_2 + x_3 &\leq 2 \quad &\text{(recurso 2)} \\ x_1 + x_2 + 3x_3 &\leq 12 \quad &\text{(recurso 3)} \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

D,I ***a***) Aplique el método símplex paso a paso para resolver el problema.
b) Identifique los precios sombra de los tres recursos y describa su significado.
C ***c***) Utilice un software basado en el método símplex para resolver el problema y elabore el informe de sensibilidad. Use esta información para identificar los precios sombra de cada recurso, el intervalo permisible para seguir óptimo de cada coeficiente de la función objetivo y el intervalo permisible para seguir factible de cada lado derecho.

4.7-6. Considere el siguiente problema.

$$\text{Maximizar} \quad Z = 5x_1 + 4x_2 - x_3 + 3x_4,$$

sujeta a

$$\begin{aligned} 3x_1 + 2x_2 - 3x_3 + x_4 &\leq 24 \quad &\text{(recurso 1)} \\ 3x_1 + 3x_2 + x_3 + 3x_4 &\leq 36 \quad &\text{(recurso 2)} \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0.$$

D,I ***a***) Aplique el método símplex paso a paso para resolver el problema.
b) Identifique los precios sombra de los tres recursos y describa su significado.
C ***c***) Utilice un software basado en el método símplex para resolver el problema y genere el informe de sensibilidad. Use esta información para identificar los precios sombra de cada recurso, el intervalo permisible para seguir óptimo de cada coeficiente de la función objetivo y el intervalo permisible para seguir factible de cada lado derecho.

4.9-1. Utilice el algoritmo de punto interior en el IOR Tutorial para resolver el modelo del problema 4.1-4. Elija $\alpha = 0.5$ del menú de opciones (Options), use $(x_1, x_2) = (0.1, 0.4)$ como la solución de prueba inicial y corra 15 iteraciones. Dibuje una gráfica de la región factible y después grafique la trayectoria de las soluciones de prueba a través de esta región factible.

4.9-2. Repita el problema 4.9-1 para el modelo del problema 4.1-5.

CASOS

Caso 4.1 Telas y moda de otoño

En el décimo piso de su edificio de oficinas, Katherine Rally observa las hordas de neoyorquinos que batallan por caminar por calles llenas de taxis y banquetas saturadas de puestos de hot dogs. En este caluroso día de julio dedica una atención especial a la moda que ostentan algunas mujeres y se pregunta qué elegirán ponerse en el otoño. Sus pensamientos no son cavilaciones aleatorias: son críticas para su trabajo pues posee y administra TrendLines, una compañía de ropa elegante para dama.

Hoy es un día en especial importante porque debe reunirse con Ted Lawson, el gerente de producción, para decidir el plan de producción del mes próximo de la línea de verano. En particular, tendrá que determinar la cantidad de cada artículo de ropa que debe producir dada la capacidad de producción de la planta, los recursos limitados y los pronósticos de demanda. La planeación precisa de la producción del mes próximo es crítica para las ventas de otoño pues los artículos que se produzcan ese mes aparecerán en las tiendas durante septiembre, y las mujeres tienden a comprar la mayor parte de su atuendo para el otoño en cuanto aparece ese mes.

Regresa a su amplio escritorio de cristal y observa los numerosos papeles que lo cubren. Su ojos vagan por los patrones de ropa diseñados hace casi seis meses, las listas de requerimientos de materiales para cada uno y las listas de pronósticos de demanda determinados por las encuestas que se aplican a los clientes en los desfiles de modas. Recuerda los días ajetreados, a veces casi de pesadilla, del diseño de la línea de otoño y su presentación en los desfiles de Nueva York, Milán y París. Al final, pagó a su equipo de seis diseñadores un total de 860 000 dólares por su trabajo en esta línea. Con el costo de contratar modelos, estilistas y artistas del maquillaje, personal para coser

y ajustar las prendas, la construcción de los escenarios, la coreografía, los ensayos para el desfile y la renta de los salones de conferencias, cada desfile le costó 2 700 000 dólares.

Se dedica a estudiar los patrones de ropa y las necesidades de materiales. Su línea de otoño consiste en ropa tanto profesional (para el trabajo) como informal. Ella determina los precios de cada prenda tomando en cuenta la calidad, el costo del material, de la mano de obra y de los maquinados, la demanda del artículo y el prestigio del nombre de la marca TrendLines.

La moda profesional para el otoño incluye:

Artículo	Requerimiento de materiales	Precio	Costos de mano de obra y maquilado
Pantalones de lana	3 yardas de lana 2 yardas de acetato para forro	$300	$160
Suéter de cashmere	1.5 yardas de cashmere	$450	$150
Blusa de seda	1.5 yardas de seda	$180	$100
Camisola de seda	0.5 yardas de seda	$120	$ 60
Falda ajustada	2 yardas de rayón 1.5 yardas de acetato para forro	$270	$120
Chaqueta de lana	2.5 yardas de lana 1.5 yardas de acetato para forro	$320	$140

La moda informal de otoño incluye:

Artículo	Requerimiento de materiales	Precio	Costos de mano de obra y maquilado
Pantalón de terciopelo	3 yardas de terciopelo 2 yardas de acetato para forro	$350	$175
Suéter de algodón	1.5 yardas de algodón	$130	$ 60
Minifalda de algodón	0.5 yardas de algodón	$ 75	$ 40
Camisa de terciopelo	1.5 yardas de terciopelo	$200	$160
Blusa de botones	1.5 yardas de rayón	$120	$ 90

Para la producción del próximo mes, ella ha ordenado 45 000 yardas de lana, 28 000 de acetato, 9 000 de cashmere, 18 000 de seda, 30 000 de rayón, 20 000 de terciopelo y 30 000 de algodón. Los precios de los materiales se presentan en la siguiente tabla:

Material	Precio por yarda
Lana	$ 9.00
Acetato	$ 1.50
Cashmere	$60.00
Seda	$13.00
Rayón	$ 2.25
Terciopelo	$12.00
Algodón	$ 2.50

Cualquier material que no se use en la producción se puede devolver al distribuidor de telas y obtener su reembolso, aunque el desperdicio no se puede devolver.

Ella sabe que la producción tanto de la blusa de seda como del suéter de algodón deja material de desperdicio. En especial, para producir una blusa de seda o un suéter de algodón, se necesitan 2 yardas de seda y 2 de algodón, respectivamente. De estas dos yardas, 1.5 se usan para la blusa o el suéter y 0.5 yardas quedan como material de desperdicio. No quiere desaprovechar este material, por lo que planea utilizar el desperdicio rectangular de seda o algodón para producir una camisola de seda o una minifalda de algodón, respectivamente. Por lo tanto, si se produce una blusa de seda, también se produce una camisola de este material. Del mismo modo, cuando se produce un suéter de algodón, también se fabrica una minifalda de algodón. Observe que es posible producir una camisola de seda sin producir una blusa del mismo material o una minifalda de algodón sin producir un suéter de algodón.

Los pronósticos de demanda indican que algunos artículos tienen una demanda limitada. En particular, dado que los pantalones y camisas de terciopelo son novedades, TrendLines ha pronosticado que puede vender sólo 5 500 pares de pantalones y 6 000 camisas. La empresa no quiere producir más de la demanda pronosticada porque una vez que pasen de moda no los podrá vender. Sin embargo, puede producir menos de lo pronosticado, ya que no se requiere que la compañía cumpla con la demanda. El suéter de cashmere también tiene una demanda limitada porque es bastante costoso, y TrendLines sabe que puede vender, como máximo, 4 000 suéteres. Las blusas de seda y las camisolas tienen demanda limitada por la idea de las mujeres de que es difícil cuidar la seda, y las proyecciones de TrendLines son que puede vender a lo más 12 000 blusas y 15 000 camisolas.

Los pronósticos de demanda también indican que los pantalones de lana, las faldas ajustadas y las chaquetas de lana tienen una gran demanda porque son artículos básicos necesarios

en todo guardarropa profesional. En especial, la demanda de los pantalones de lana es de 7 000 pares y la de las chaquetas de 5 000 unidades. Katherine desea cumplir con al menos 60% de la demanda de estos dos artículos para mantener la lealtad de su base de clientes y no perderlos en el futuro. Aunque la demanda de faldas no se puede estimar, Katherine siente que debe producir al menos 2 800 de ellas.

a) Ted intenta convencer a Katherine de no producir camisas de terciopelo pues la demanda de esta moda novedosa es baja. Afirma que sólo es responsable de 500 000 dólares de los costos de diseño y otros costos fijos. La contribución neta (precio del artículo-costos de materiales-costo de mano de obra) cuando se venda la novedad debe cubrir estos costos fijos. Cada camisa de terciopelo genera una contribución neta de 22 dólares. Él afirma que dada la contribución neta, aun si se satisface la demanda máxima, no dejará ganancias. ¿Qué piensa del argumento de Ted?

b) Formule y resuelva un problema de programación lineal para maximizar la ganancia dadas las restricciones de producción, recursos y demanda.

Antes de tomar una decisión final, Katherine planea explorar las siguientes preguntas independientes, excepto cuando se indique otra cosa.

c) El distribuidor de textiles informa a Katherine que no puede recibirle el terciopelo sobrante porque los pronósticos de demanda muestran que la demanda de esta tela disminuirá en el futuro. En consecuencia, Katherine no obtendrá el reembolso por el terciopelo. ¿En qué cambia este hecho el plan de producción?

d) ¿Cuál es una explicación económica intuitiva de la diferencia entre las soluciones que se encontraron en los incisos *b*) y *c*)?

e) El personal de costura encuentra dificultades para coser los forros de las mangas de los sacos de lana pues el patrón tiene una forma extraña y el pesado material de lana es difícil de cortar y coser. El incremento de tiempo para coser un saco de lana aumenta en 80 dólares los costos de mano de obra y maquinado por cada saco. Dado este nuevo costo, ¿cuántas prendas de cada tipo debe producir TrendLines para maximizar la ganancia?

f) El distribuidor de textiles informa a Katherine que como otro cliente canceló su orden, ella puede obtener 10 000 yardas adicionales de acetato. ¿Cuántas prendas de cada tipo debe producir TrendLines para maximizar la ganancia?

g) TrendLines supone que puede vender todas las prendas que no se vendan en septiembre y octubre en una gran barata en noviembre a 60% de su precio original. Por lo tanto, en esa oportunidad puede vender cantidades ilimitadas de todos los artículos. (Los límites superiores mencionados se refieren sólo a las ventas durante septiembre y octubre.) ¿Cuál debe ser el nuevo plan de producción para maximizar la ganancia?

RESUMEN DE LOS CASOS ADICIONALES EN EL SITIO EN INTERNET DEL LIBRO (www.mhhe.com/hillier)

Caso 4.2 Nuevas fronteras

AmeriBank comenzará muy pronto a ofrecer a sus clientes el servicio de banca en red. Con la intención de guiar su planeación para los servicios que proporcionará a través de internet, se realizará una encuesta entre cuatro diferentes grupos de edad y tres tipos de comunidades. AmeriBank impone ciertas restricciones acerca de la profundidad con que debe ser encuestado cada grupo de edad y tipo de comunidad. Se requiere la programación lineal para desarrollar un plan para la encuesta que minimice su costo total al mismo tiempo que satisfaga todas las restricciones bajo varios escenarios diferentes.

Caso 4.3 Asignación de estudiantes a escuelas

La oficina escolar de Springfield decidió cerrar una de sus escuelas de educación media; por ello necesita reasignar a todos los estudiantes de nivel medio para el próximo año en las tres escuelas restantes. Muchos de ellos serán transportados en autobús, por lo que la minimización de los costos de transporte escolar es uno de los objetivos. Otro es minimizar los elementos de inconveniencia y seguridad de los estudiantes que deberán trasladarse a la escuela a pie o en bicicleta. Dadas las capacidades de las tres escuelas, así como la necesidad de encontrar un punto cercano al balance para el número de estudiantes en los tres grados de cada escuela, ¿cómo puede utilizarse la programación lineal para determinar cuántos estudiantes de cada una de las seis áreas residenciales de la ciudad deben ser asignados a cada escuela? ¿Qué pasaría si la totalidad de cada área residencial debiera ser asignada a la misma escuela? (Este caso continuará en las casos 6.3 y 11.4.)

CAPÍTULO 5

Teoría del método símplex

En el capítulo 4 se presentó el mecanismo básico del método símplex. Ahora se profundizará un poco en este algoritmo al examinar parte de la teoría en que se apoya. En esta primera sección se estudian a detalle las propiedades algebraicas y geométricas generales que constituyen el fundamento del método símplex. Después se describe la *forma matricial* del método símplex, que simplifica en gran medida el procedimiento para realizarlo en computadora. En seguida, se utiliza esta forma matricial para presentar la idea fundamental sobre la propiedad del método símplex que permite deducir de qué manera los cambios que se hacen en el modelo original repercuten en la tabla símplex final. Esta idea proporcionará la clave para los temas importantes del capítulo 6 (teoría de dualidad y análisis de sensibilidad). Este capítulo concluye con la presentación del *método símplex revisado*, que simplifica aún más la forma matricial del método símplex. Por lo general, los programas de cómputo comerciales del método símplex están basados en el método símplex revisado.

5.1 FUNDAMENTOS DEL MÉTODO SÍMPLEX

En la sección 4.1 se introdujo el concepto de *soluciones factibles en un vértice* (*FEV*) y la función clave que desempeñan en el método símplex. Estos conceptos geométricos se relacionaron con el álgebra del método símplex de las secciones 4.2 y 4.3. Sin embargo, todo esto se hizo en el contexto del problema de la Wyndor Glass Co., que tiene sólo *dos variables de decisión* y, por lo mismo, tiene una interpretación geométrica directa. ¿Cómo pueden generalizarse estos conceptos a dimensiones mayores cuando se manejan problemas más grandes? La respuesta se dará en esta sección.

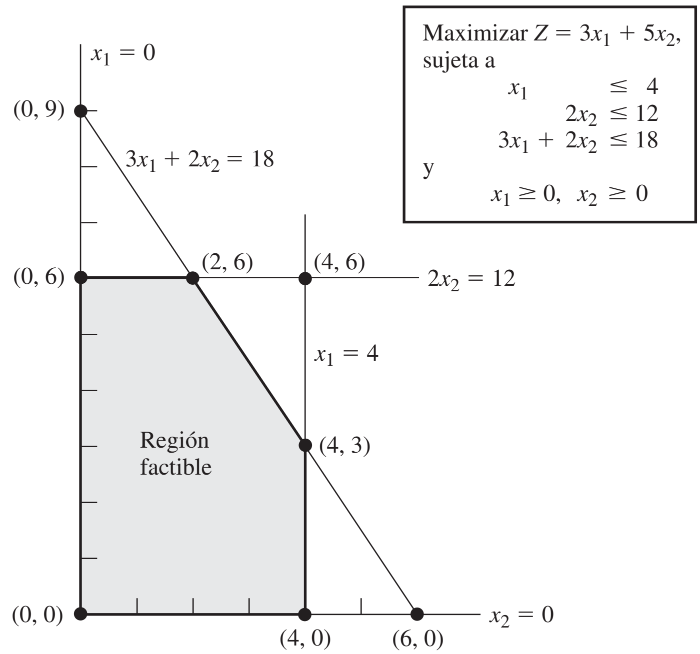
Para comenzar se introducirá parte de la terminología básica de cualquier problema de programación lineal con n variables de decisión. Mientras se desarrolla esta tarea puede ser útil que el lector consulte la figura 5.1 (que es una repetición de la figura 4.1) para interpretar estas definiciones en dos dimensiones ($n = 2$).

Terminología

Puede entenderse de manera intuitiva que las soluciones óptimas de cualquier problema de programación lineal deben estar sobre la frontera de la región factible y, de hecho, ésta es una propiedad general. Como la frontera es un concepto geométrico, las definiciones iniciales aclaran cómo se puede identificar en forma algebraica la frontera de la región factible.

> La **ecuación de la frontera de restricción** de cualquier restricción se obtiene al sustituir su signo $\leq$, $=$ o $\geq$ por un signo $=$.

En consecuencia, la forma de la ecuación de una frontera de restricción es $a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i$ para las restricciones funcionales y $x_j = 0$ en el caso de las de no negatividad. Estas ecuaciones definen una figura geométrica "plana" (llamada **hiperplano**) en un espacio n dimensional, análoga a la recta en el espacio bidimensional y al plano en el espacio tridimensional. Este hiperplano forma la **frontera de restricción** de la restricción correspondiente. Cuando la restricción



■ **FIGURA 5.1**
Fronteras de restricción, ecuaciones de las fronteras de restricción y soluciones en los vértices del problema de la Wyndor Glass Co.

tiene signo ≤ o ≥, esta *frontera de restricción* separa los puntos que satisfacen la restricción (todos los puntos que se encuentran en un lado e incluyen a la frontera de restricción) de los puntos que la violan (todos aquellos que se encuentran del otro lado de la frontera de restricción). Cuando la restricción tiene un signo =, sólo los puntos sobre la frontera de restricción la satisfacen.

Por ejemplo, el problema de la Wyndor Glass Co. tiene cinco restricciones (tres funcionales y dos de no negatividad), de manera que tiene las cinco *ecuaciones de frontera de restricción* que se muestran en la figura 5.1. Como $n = 2$, los hiperplanos definidos por estas ecuaciones de frontera de restricción son simples rectas. Por lo tanto, las fronteras de restricción de las cinco restricciones son las cinco rectas que se muestran en la figura 5.1.

La **frontera** de la región factible consiste en aquellas soluciones factibles que satisfacen una o más de las ecuaciones de frontera de las restricciones.

En el sentido geométrico, cualquier punto sobre la frontera de la región factible se encuentra sobre uno o más de los hiperplanos definidos por las ecuaciones de frontera de restricción respectivas. En la figura 5.1, la frontera consiste en los cinco segmentos de recta más oscuros.

En seguida se da una definición general de *solución FEV* en el espacio de *n* dimensiones.

Una **solución factible en un vértice** (**FEV**) es una solución factible que no se encuentra en *cualquier* segmento rectilíneo[1] que conecta a *otras* dos soluciones factibles.

Como está implícito en esta definición, una solución factible que *está* sobre un segmento rectilíneo que conecta a otras dos soluciones factibles *no* es una solución FEV. Para ilustrar el caso en que $n = 2$, considere la figura 5.1. El punto (2, 3) *no* es una solución FEV, puesto que se encuentra en varios de estos segmentos, por ejemplo, en el segmento de recta que conecta los puntos (0, 3) y (4, 3). De igual manera, (0, 3) *no* es una solución FEV, porque se encuentra sobre el segmento de recta que conecta a (0, 0) con (0, 6). Sin embargo, (0, 0) es una solución FEV, porque es imposible hallar *otras* dos soluciones factibles que se encuentren en lados completamente opuestos de (0, 0). (Intente encontrarla.)

Cuando el número *n* de variables de decisión es mayor que 2 o 3, esta definición de *soluciones FEV* no es muy conveniente para identificarlas. Por lo tanto, una interpretación algebraica de estas soluciones resultará más útil. En el ejemplo de la Wyndor Glass Co., cada solución FEV de la figura 5.1 está en la intersección de dos rectas de restricción ($n = 2$), es decir, es una *solución simultánea*

[1] En el apéndice 2 se presenta una expresión algebraica de un segmento de recta.

de un sistema de dos ecuaciones de frontera de restricción. Esta situación se resume en la tabla 5.1, en la que las **ecuaciones de definición** se refieren a las ecuaciones de frontera de restricción que conducen a, o definen, las soluciones FEV indicadas.

> Para cualquier problema de programación lineal con n variables de decisión, cada solución FEV se encuentra en la intersección de n fronteras de restricción; esto es, se trata de una *solución simultánea* de un sistema de n ecuaciones de frontera de restricción.

No obstante, esto no quiere decir que *todo* conjunto de n ecuaciones de frontera de restricción que se elija entre las $n + m$ restricciones (n restricciones de no negatividad y m restricciones funcionales) conduce a una solución factible en un vértice. En particular, la solución simultánea de un sistema de ecuaciones tal, puede violar una o más de las otras m restricciones no seleccionadas, en cuyo caso se trata de una solución *no factible* en un vértice. El ejemplo tiene tres soluciones de este tipo, como se resume en la tabla 5.2. (Verifique por qué son no factibles.)

Más aún, un sistema de n ecuaciones de frontera de restricción puede no tener solución. Esto ocurre dos veces en el ejemplo con los pares de ecuaciones: 1) $x_1 = 0$ y $x_1 = 4$, y 2) $x^2 = 0$ y $2x_2 = 12$. Tales sistemas no son de interés en el contexto de este libro.

La última posibilidad (que no ocurre en el ejemplo) es que un sistema de n ecuaciones de frontera de restricción tenga soluciones múltiples debido a ecuaciones redundantes. Tampoco es necesario preocuparse por un hecho de este tipo, puesto que el método símplex evita las dificultades del caso.

Debemos también hacer mención que es posible que más de un sistema con n ecuaciones de frontera de restricción nos arroje la misma solución CP. Por ejemplo, si la restricción $x_1 \leq 4$ en el caso del problema de Wyndor Glass Co. se fuera a reemplazar por $x_1 \leq 2$, observe en la figura 5.1 cómo se puede deducir la solución (2, 6) de cualquiera de los tres pares de ecuaciones de frontera de restricción. (Éste es un ejemplo de la *degeneración* que se estudió en un contexto diferente en la sección 4.5.)

Con el fin de resumir el ejemplo, con cinco restricciones y dos variables, existen 10 pares de ecuaciones de frontera de restricción. Cinco de ellos definieron las ecuaciones de las soluciones FEV (tabla 5.1), tres definieron las ecuaciones de las soluciones en vértice no factibles (tabla 5.2) y los dos pares del final no tuvieron solución.

Soluciones FEV adyacentes

En la sección 4.1 se presentaron las soluciones FEV adyacentes y su función en la solución de problemas de programación lineal. Ahora, se profundizará sobre esto.

Recuerde que en el capítulo 4 (cuando se ignoran las variables de holgura, de exceso y artificiales) cada iteración del método símplex se mueve de la solución FEV actual a una *adyacente*. ¿Cuál es la *trayectoria* que sigue este proceso? ¿Qué significa en realidad una solución FEV *adyacente*? Primero se contestará a estas preguntas desde el punto de vista geométrico y después se darán las interpretaciones algebraicas.

TABLA 5.1 Ecuaciones de definición de cada solución FEV del problema de la Wyndor Glass Co.

Solución FEV	Ecuaciones de definición
(0, 0)	$x_1 = 0$ $x_2 = 0$
(0, 6)	$x_1 = 0$ $2x_2 = 12$
(2, 6)	$2x_2 = 12$ $3x_1 + 2x_2 = 18$
(4, 3)	$3x_1 + 2x_2 = 18$ $x_1 = 4$
(4, 0)	$x_1 = 4$ $x_2 = 0$

TABLA 5.2 Ecuaciones de definición de cada solución no factible en un vértice del problema de la Wyndor Glass Co.

Solución no factible en un vértice	Ecuaciones de definición
(0, 9)	$x_1 = 0$ $3x_1 + 2x_2 = 18$
(4, 6)	$2x_2 = 12$ $x_1 = 4$
(6, 0)	$3x_1 + 2x_2 = 18$ $x_2 = 0$

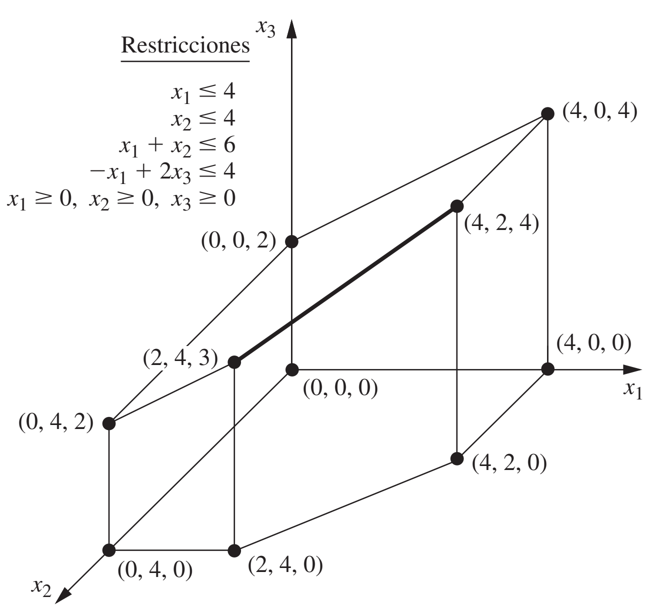
La respuesta a estas preguntas es sencilla cuando $n = 2$. En este caso, la *frontera* de la región factible consiste en varios *segmentos de recta* conectados de manera que forman un *polígono*, como lo muestran, en la figura 5.1, los cinco segmentos más oscuros. Estos segmentos de recta se conocen como *aristas* de la región factible. De cada solución FEV se originan dos de estas aristas que llevan a una solución FEV adyacente en el otro punto terminal. (Observe en la figura 5.1 que cada solución FEV tiene dos soluciones adyacentes.) Cada iteración sigue una trayectoria a lo largo de estas aristas al moverse de un punto terminal a otro. En la figura 5.1 la primera iteración se mueve a lo largo de la arista que va de (0, 0) a (0, 6), y luego la siguiente iteración se mueve por la arista que va de (0, 6) a (2, 6). Como se puede ver en la tabla 5.1, cada uno de estos movimientos a una solución FEV adyacente significa sólo un cambio en el conjunto de ecuaciones de definición (fronteras de restricción sobre las que se encuentra la solución).

Cuando $n = 3$, las respuestas son un poco más complejas. Para ayudar a visualizar lo que ocurre, la figura 5.2 muestra un dibujo en tres dimensiones de una región factible representativa para este caso, en la que los puntos son las soluciones FEV. Esta región factible es un *poliedro* en lugar del polígono que se tenía con $n = 2$ (figura 5.1), ya que las fronteras de restricción son ahora planos y no rectas. Las caras del poliedro forman la *frontera* de la región factible, donde cada cara es la porción de la frontera de restricción que también satisface las otras restricciones. Observe que cada solución FEV se encuentra en la intersección de tres fronteras de restricción (a veces incluye parte de las fronteras de restricción $x_1 = 0$, $x_2 = 0$ y $x_3 = 0$ para las restricciones de no negatividad) y que la solución también satisface las otras restricciones. Las intersecciones que no satisfacen una o más de las otras restricciones conducen a soluciones *no factibles* en un vértice.

El segmento de recta más oscuro de la figura 5.2 indica la trayectoria que sigue el método símplex en una iteración normal. El punto (2, 4, 3) es la solución FEV *actual* que se usa para iniciar una iteración y el punto (4, 2, 4) será la nueva solución FEV al término de la iteración. El punto (2, 4, 3) se encuentra en la intersección de las fronteras de restricción $x_2 = 4$, $x_1 + x_2 = 6$ y $-x_1 + 2x_3 = 4$, por lo que estas tres ecuaciones son las ecuaciones de definición para esta solución FEV. Si se eliminara la *ecuación de definición* $x_2 = 4$, la intersección de las otras dos fronteras de restricción (planos) formarían una recta. Un segmento de esta recta, que aparece en la figura 5.2 como el segmento más oscuro que va desde (2, 4, 3) hasta (4, 2, 4), está sobre la frontera de la región factible, mientras que el resto de la recta es no factible. Este segmento de recta se llama arista de la región factible y sus puntos terminales (2, 4, 3) y (4, 2, 4) son soluciones FEV adyacentes.

De esta manera, para $n = 3$, todas las *aristas* de la región factible se forman como el segmento factible de la recta que está en la intersección de dos fronteras de restricción, y los dos puntos terminales de una arista son soluciones FEV *adyacentes*. En la figura 5.2 hay 15 aristas de la región factible y, por lo tanto, hay 15 pares de soluciones FEV adyacentes. Para la solución FEV (2, 4, 3) actual,

FIGURA 5.2
Región factible y soluciones FEV de un problema de programación lineal de tres variables.



hay tres formas de eliminar una de sus tres ecuaciones de definición para obtener la intersección de las otras dos fronteras de restricción, así que hay tres aristas que se originan de (2, 4, 3). Estas aristas conducen a (4, 2, 4), (0, 4, 2) y (2, 4, 0), y son las soluciones FEV adyacentes a (2, 4, 3).

En la siguiente iteración, el método símplex elige una de estas tres aristas; por ejemplo, el segmento de recta más oscuro en la figura 5.2, y se mueve a lo largo de él cada vez más lejos de (2, 4, 3) hasta que llega a la primera frontera de restricción nueva, $x_1 = 4$, en la otra punta. [No se puede continuar por esta recta hasta la siguiente frontera de restricción, $x_2 = 0$, porque se llegaría a una solución no factible en un vértice: (6, 0, 5).] La intersección de esta primera frontera de restricción nueva con las dos fronteras de restricción que forman la arista conduce a la *nueva* solución FEV, (4, 2, 4).

Cuando $n > 3$, estos mismos conceptos se generalizan a dimensiones mayores, sólo que las fronteras de restricción son ahora *hiperplanos* en lugar de planos. En resumen:

> Considere cualquier problema de programación lineal con n variables de decisión y una región factible acotada. Una solución FEV se encuentra en la intersección de n fronteras de restricción (y satisface también las otras restricciones). Una **arista** de la región factible es un segmento de recta factible que está en la intersección de $n - 1$ fronteras de restricción, donde cada punto terminal se encuentra en una frontera de restricción adicional (por lo que estos puntos terminales son soluciones FEV). Dos soluciones factibles en un vértice son **adyacentes** si el segmento de recta que las conecta es una arista de la región factible. De cada solución FEV se originan n aristas, cada una de las cuales conduce a una de las n soluciones FEV adyacentes. Cada iteración del método símplex se mueve de la solución FEV actual a una adyacente a lo largo de una de estas n aristas.

Al cambiar del punto de vista geométrico al algebraico, la *intersección de las fronteras de restricción* cambia a una *solución simultánea de las ecuaciones de frontera de restricción*. Las n ecuaciones de frontera de restricción que conducen a (o definen) una solución FEV son sus ecuaciones de definición; al eliminar una de estas ecuaciones se obtiene una recta cuyo segmento factible es una arista de la región factible.

En seguida se analizarán algunas propiedades importantes de las soluciones FEV y luego se describirán las implicaciones de todos estos conceptos al interpretar el método símplex. Sin embargo, ahora que se tiene fresco el resumen anterior, se expondrá un panorama general de estas implicaciones. Cuando el método símplex elige una variable básica entrante, la interpretación geométrica es que está eligiendo una de las aristas que se originan de la solución FEV actual para trasladarse por ella. El aumento del valor de esta variable a partir de cero (y al mismo tiempo el cambio en los valores de las otras variables básicas según sea necesario) corresponde a moverse a lo largo de esta arista. El hecho de que una de estas variables básicas (la variable básica que sale) disminuya su valor hasta que llega a cero corresponde a llegar a la primera frontera de restricción nueva en el otro punto terminal de esta arista de la región factible.

Propiedades de las soluciones FEV

En este momento la atención se enfocará en tres propiedades fundamentales de las soluciones FEV que se cumplen en el caso de *cualquier* problema de programación lineal que tiene soluciones factibles y una región factible acotada.

> **Propiedad 1:** *a*) Si el problema tiene sólo una solución óptima, ésta debe ser una solución FEV. *b*) Si el problema tiene soluciones óptimas múltiples (y una región factible acotada), entonces al menos dos deben ser soluciones FEV adyacentes.

La propiedad 1 es un concepto bastante intuitivo desde el punto de vista geométrico. Primero, considere el caso *a*), que se ilustra con el problema de la Wyndor Glass Co. (vea la figura 5.1), donde la única solución óptima (2, 6) es, sin duda, una solución FEV. Observe que no hay nada especial en el ejemplo que conduzca a este resultado. Para cualquier problema con una sola solución óptima, siempre es posible elevar cada vez más la recta (hiperplano) de la función objetivo hasta que toque sólo un punto (la solución óptima) en un vértice de la región factible.

Ahora se dará una demostración algebraica de este caso.

> **Demostración del caso *a*) de la propiedad 1:** Se desarrolla una *demostración por contradicción* que se basa en el supuesto de que existe sólo una solución óptima y que *no* es

una solución FEV. Después se demuestra que este supuesto lleva a una contradicción y, por lo tanto, no puede ser cierta. (La solución que se supone óptima se denota por $\mathbf{x}^*$ y el valor correspondiente de la función objetivo por Z^*.)

Recuerde la definición de la *solución FEV* (una solución factible que no está en ningún segmento que conecte a otras dos soluciones factibles). Como se ha supuesto que la solución óptima $\mathbf{x}^*$ no es una solución FEV, se desprende que deben existir otras dos soluciones factibles tales que el segmento de recta que las une contiene la solución óptima. Sean $\mathbf{x}'$ y $\mathbf{x}''$ las otras dos soluciones factibles y sean Z_1 y Z_2 los valores respectivos de la función objetivo. Igual que para cualquier otro punto sobre el segmento de recta que conecta a $\mathbf{x}'$ y $\mathbf{x}''$,

$$\mathbf{x}^* = \alpha\mathbf{x}'' + (1 - \alpha)\mathbf{x}'$$

para algún valor de α tal que $0 < \alpha < 1$. Entonces, puesto que los coeficientes de las variables son idénticos para Z^*, Z_1 y Z_2, se puede deducir que

$$Z^* = \alpha Z_2 + (1 - \alpha)Z_1.$$

Como las ponderaciones α y $1 - \alpha$ suman 1, las únicas posibilidades que se tienen al comparar Z^*, Z_1 y Z_2 son: 1) $Z^* = Z_1 = Z_2$, 2) $Z_1 < Z^* < Z_2$ y 3) $Z_1 > Z^* > Z_2$. La primera posibilidad implica que $\mathbf{x}'$ y $\mathbf{x}''$ también son óptimas, lo que contradice la suposición de que existe sólo una solución óptima. Las otras dos posibilidades contradicen la suposición de que $\mathbf{x}^*$ (no una solución FEV) es óptima. En conclusión, es imposible tener una solución óptima que no sea una solución FEV.

Ahora considere el caso *b*) que se demostró en la sección 3.2 bajo la definición de *solución óptima* al cambiar la función objetivo del ejemplo a $Z = 3x_1 + 2x_2$ (vea la figura 3.5 en la sección 3.2). Lo que ocurrió en el procedimiento gráfico es que la recta que representa a la función objetivo se mueve hacia arriba hasta que contiene el segmento de recta que conecta las dos soluciones FEV (2, 6) y (4, 3). Lo mismo pasa en dimensiones mayores, excepto que la función objetivo es un *hiperplano* que se mueve hasta que contiene el o los segmentos que conectan dos (o más) soluciones FEV adyacentes. Como consecuencia, es posible obtener *todas* las soluciones óptimas como promedios ponderados de soluciones FEV óptimas. (Esta situación se describe con más detalle en los problemas 4.5-5 y 4.5-6.)

El significado real de la propiedad 1 es que simplifica mucho la búsqueda de una solución óptima, ya que ahora sólo tienen que tomarse en cuenta las soluciones FEV. La propiedad 2 pone de relieve la magnitud de esta simplificación.

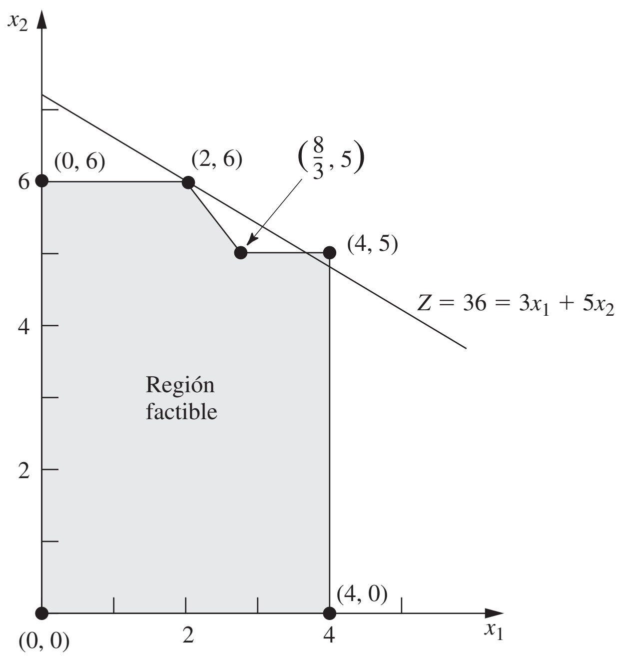
Propiedad 2: Existe sólo un número *finito* de soluciones FEV.

Esta propiedad sin duda se cumple en las figuras 5.1 y 5.2, donde nada más se tienen 5 y 10 soluciones FEV, respectivamente. Para ver por qué en general este número es finito, recuerde que cada solución factible en un vértice es la solución simultánea de un sistema de n ecuaciones elegidas entre $m + n$ ecuaciones de frontera de restricción. El número de combinaciones de las $m + n$ ecuaciones tomadas n a la vez es

$$\binom{m+n}{n} = \frac{(m+n)!}{m!n!},$$

que es un número finito. Este número, a su vez, es la *cota superior* del número de soluciones FEV. En la figura 5.1, $m = 3$ y $n = 2$, de manera que existen 10 sistemas diferentes de dos ecuaciones, pero sólo de la mitad de ellos se obtienen soluciones FEV. En la figura 5.2, $m = 4$ y $n = 3$, de manera que hay 35 sistemas diferentes de tres ecuaciones, pero sólo 10 conducen a soluciones FEV.

La propiedad 2 sugiere que, en principio, puede obtenerse una solución óptima mediante la enumeración exhaustiva; esto es, se pueden encontrar y comparar todas las soluciones FEV, que son un número finito. Desafortunadamente existen números finitos que (para propósitos prácticos) bien podrían considerarse infinitos. Por ejemplo, un problema de programación lineal bastante pequeño con $m = 50$ y $n = 50$ tendría $100!/(50!)^2$, aproximadamente 10^{29} sistemas de ecuaciones que resolver. En contraste, el método símplex necesitaría examinar nada más alrededor de 100 soluciones FEV para resolver un problema de este tamaño. Este ahorro tan grande se puede obtener gracias a la prueba de optimalidad dada en la sección 4.1 y que se establece aquí como la propiedad 3.



■ **FIGURA 5.3**
Modificación del problema de la Wyndor Glass Co., que viola tanto la programación lineal como la propiedad 3 de las soluciones FEV en programación lineal.

Propiedad 3: Si una solución FEV no tiene soluciones FEV *adyacentes* que sean *mejores* que ella (en términos del valor de *Z*), entonces no existen soluciones FEV que sean *mejores* en cualquier otra parte. Por lo tanto, se garantiza que tal solución FEV es una solución *óptima* (por la propiedad 1), si se supone que el problema tiene al menos una solución óptima (lo que se garantiza si el problema tiene soluciones factibles y una región factible acotada).

Para ilustrar la propiedad 3 considere la figura 5.1 del ejemplo de la Wyndor Glass Co. Las soluciones FEV (0, 6) y (4, 3) son adyacentes a la solución FEV (2, 6) y ninguna de las dos tiene un valor mejor de *Z*. Esto implica que ninguna de las otras soluciones FEV, (0, 0) y (0, 4), pueden ser mejores que (2, 6), por lo cual (2, 6) debe ser óptima.

Por el contrario, en la figura 5.3 se muestra una región factible que *nunca* se puede presentar en un problema de programación lineal [puesto que la continuación de la recta de la frontera de restricción que pasa por $(\frac{8}{3}, 5)$ cortaría una parte de esta región], pero esto viola la propiedad 3. El problema que se muestra es idéntico al de la Wyndor Glass Co. (incluso con la misma función objetivo), *excepto* que se aumentó la región factible hacia la derecha de $(\frac{8}{3}, 5)$. En consecuencia, las soluciones FEV adyacentes a (2, 6) son ahora (0, 6) y $(\frac{8}{3}, 5)$, y de nuevo ninguna de las dos es mejor que (2, 6). Sin embargo, ahora otra solución FEV, (4, 5), es mejor que (2, 6), lo que viola la propiedad 3. La razón es que la frontera de la región factible va hacia abajo desde (2, 6) hacia $(\frac{8}{3}, 5)$ y después "dobla hacia afuera" hasta (4, 5), más allá de la recta de la función objetivo que pasa por (2, 6).

El punto clave es que el tipo de situación que se ilustra en la figura 5.3 no se puede presentar en programación lineal. La región factible de esta figura implica que las restricciones $2x_2 \leq 12$ y $3x_1 + 2x_2 \leq 18$ se cumplen para $0 \leq x_1 \leq \frac{8}{3}$. Sin embargo, bajo la condición de que $\frac{8}{3} \leq x_1 \leq 4$, la restricción $3x_1 + 2x_2 \leq 18$ se elimina y es reemplazada por $x_2 \leq 5$. Este tipo de "restricciones condicionales" no están permitidas en programación lineal.

La razón básica de que la propiedad 3 se cumpla es que en cualquier problema de programación lineal, la región factible siempre tiene la propiedad de ser un *conjunto convexo*,[2] según se de-

[2] Si ya se tiene conocimiento de los conjuntos convexos es posible notar que el conjunto de soluciones que satisfacen cualquier restricción de programación lineal (ya sea una restricción de igualdad o desigualdad) es un conjunto convexo. La región factible de cualquier problema de programación lineal es la *intersección* de los conjuntos de soluciones que satisfacen sus restricciones individuales. Como la intersección de conjuntos convexos es otro conjunto convexo, esta región factible necesariamente también lo es.

fine en el apéndice 2 y se ilustra en varias figuras. En un problema de dos variables, esta propiedad de convexidad significa que el *ángulo* dentro de la región factible en *todas* las soluciones FEV es menor que 180°, situación que se ilustra en la figura 5.1, donde los ángulos en (0, 0), (0, 6) y (4, 0) son de 90° y los que se encuentran en (2, 6) y (4, 3) tienen entre 90 y 180°. Por el contrario, la región factible de la figura 5.3 *no* es un conjunto convexo debido a que el ángulo en ($\frac{8}{3}$, 5) es mayor que 180°. Éste es el tipo de "doblez hacia afuera" con un ángulo mayor que 180° que no puede ocurrir en programación lineal. En problemas de *n* dimensiones se sigue aplicando este concepto intuitivo de "nunca doblar hacia afuera" (una propiedad fundamental de un conjunto convexo).

Para aclarar el significado de región factible convexa, considere el hiperplano de la función objetivo que pasa por una solución FEV y que es igual o mejor que todas las soluciones FEV adyacentes. [En el ejemplo original de la Wyndor Glass Co. este hiperplano es la recta de la función objetivo que pasa por (2, 6).] Todas estas soluciones adyacentes [(0, 6) y (4, 3) en el ejemplo] deben estar ya sea en el hiperplano o en el lado no favorable de éste (de acuerdo con el valor de *Z*). Cuando la región factible es convexa, su frontera no se puede "doblar hacia afuera" más allá de una solución FEV adyacente para dar otra solución FEV que se encuentre en el lado favorable del hiperplano, para que la propiedad 3 se cumpla.

Extensiones a la forma aumentada del problema

En cualquier problema de programación lineal en nuestra forma estándar (inclusive las restricciones funcionales de la forma ≤), la apariencia de las restricciones funcionales después de introducir variables de holgura es la siguiente:

$$\begin{array}{llcl}
(1) & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + x_{n+1} & & = b_1 \\
(2) & a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \quad + x_{n+2} & & = b_2 \\
 & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & & \\
(m) & a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & + x_{n+m} & = b_m,
\end{array}$$

donde $x_{n+1}, x_{n+2}, \ldots, x_{n+m}$ son las variables de holgura. En la sección 4.6 se describió cómo puede obtenerse esta misma forma (la forma apropiada de eliminación gaussiana) en otros problemas de programación lineal, mediante la introducción de variables artificiales, etc. En consecuencia, las soluciones originales ($x_1, x_2, \ldots, x_n$) quedan aumentadas con los valores correspondientes de las variables de holgura o artificiales ($x_{n+1}, x_{n+2}, \ldots, x_{n+m}$) y quizá también algunas variables de exceso. A partir de este aumento, en la sección 4.2 se definieron las **soluciones básicas** como *soluciones en un vértice aumentadas* y las **soluciones básicas factibles (soluciones BF)** como *soluciones FEV aumentadas*. En consecuencia, las tres propiedades anteriores de las soluciones FEV también se cumplen para las soluciones BF.

Ahora deben aclararse las relaciones algebraicas entre las soluciones básicas y las soluciones en los vértices. Recuerde que cada solución en un vértice es la solución simultánea de un sistema de *n* ecuaciones de frontera, a las que se dio el nombre de *ecuaciones de definición*. La pregunta clave es: ¿cómo puede distinguirse cuando una ecuación de frontera en particular es una de las ecuaciones de definición si el problema se encuentra en la forma aumentada? Por fortuna, la respuesta es sencilla. Cada restricción tiene una **variable indicativa** que señala claramente (según su valor sea cero o no) cuándo la solución actual satisface la ecuación de frontera de esa restricción. En la tabla 5.3 aparece un resumen. En el caso del tipo de restricción de cada renglón de la tabla, observe que la ecuación de frontera de restricción correspondiente (cuarta columna) se satisface si y sólo si la variable indicativa de esta restricción (quinta columna) es igual a cero. En el último renglón (restricción funcional de la forma ≥), de hecho, la variable indicativa $\bar{x}_{n+i} - x_{s_i}$ es realmente la diferencia entre la variable artificial $\bar{x}_{n+i}$ y la variable de exceso x_{s_i}.

Así, siempre que la ecuación de frontera de restricción sea una de las ecuaciones de definición de una solución en un vértice, su variable indicativa tiene valor de cero en la forma aumentada del problema. Cada una de estas variables indicativas se llama *variable no básica* de la solución básica correspondiente. En seguida se resumen las conclusiones y la terminología (que se introdujo en la sección 4.2).

> Cada **solución básica** tiene *m* **variables básicas**, y el resto son **variables no básicas** iguales a cero. (El número de variables no básicas es igual a *n* más el número de variables de exceso.) Los valores de las **variables básicas** constituyen la solución simultánea del

■ **TABLA 5.3 Variables indicativas de las ecuaciones de frontera de restricción***

Tipo de la restricción	Forma de la restricción	Restricción en la forma aumentada	Ecuación de frontera restricción	Variable indicativa
No negatividad	$x_j \geq 0$	$x_j \geq 0$	$x_j = 0$	x_j
Funcional (≤)	$\sum_{j=1}^{n} a_{ij}x_j \leq b_i$	$\sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i$	$\sum_{j=1}^{n} a_{ij}x_j = b_i$	x_{n+i}
Funcional (=)	$\sum_{j=1}^{n} a_{ij}x_j = b_i$	$\sum_{j=1}^{n} a_{ij}x_j + \bar{x}_{n+i} = b_i$	$\sum_{j=1}^{n} a_{ij}x_j = b_i$	$\bar{x}_{n+i}$
Funcional (≥)	$\sum_{j=1}^{n} a_{ij}x_j \geq b_i$	$\sum_{j=1}^{n} a_{ij}x_j + \bar{x}_{n+i} - x_{s_i} = b_i$	$\sum_{j=1}^{n} a_{ij}x_j = b_i$	$\bar{x}_{n+i} - x_{s_i}$

* Variable indicativa $= 0 \Rightarrow$ se satisface la ecuación de frontera de restricción.
Variable indicativa $\neq 0 \Rightarrow$ se viola la ecuación de frontera de restricción.

sistema de m ecuaciones del problema en la forma aumentada (después de igualar a cero las variables no básicas). Esta solución básica es la solución en el vértice aumentada cuyas n ecuaciones de definición son las indicadas por las variables no básicas. En particular, siempre que una variable indicativa de la quinta columna de la tabla 5.3 sea una variable no básica, la ecuación de frontera de restricción en la cuarta columna es una ecuación de definición para la solución en el vértice. (Para restricciones funcionales de la forma ≥, al menos una de las dos variables suplementarias $\bar{x}_{n+i}$ y x_{s_i} es siempre una variable no básica, pero la ecuación de frontera de restricción se convierte en una ecuación de definición sólo si *ambas* variables son no básicas.)

Ahora considere las soluciones básicas *factibles*. Se puede observar que los únicos requisitos para que una solución sea factible en la forma aumentada del problema son que satisfaga el sistema de ecuaciones y que *todas* las variables sean *no negativas*.

Una **solución BF** es una solución básica en la que las m variables básicas son no negativas (≥ 0). Se dice que una solución BF es **degenerada** si cualquiera de estas m variables es igual a cero.

En consecuencia, es posible que una variable sea igual a cero y no sea una variable no básica de la solución BF actual. (Este caso corresponde a una solución FEV que satisface otra frontera de restricción además de sus n ecuaciones de definición.) Por lo tanto, es necesario saber con exactitud cuál es el conjunto actual de variables no básicas (o el conjunto actual de variables básicas) y no confiar en sus valores iguales a cero.

Ya se hizo notar que no todo sistema de n ecuaciones de frontera conduce a una solución en un vértice, ya sea porque el sistema no tiene soluciones o porque tiene soluciones múltiples. Por razones análogas, no todo conjunto de n variables no básicas conduce a una solución básica. Sin embargo, el método símplex evita estos casos.

Para ejemplificar estas definiciones considere una vez más el ejemplo de la Wyndor Glass Co. Sus ecuaciones de frontera y las variables indicativas se muestran en la tabla 5.4.

■ **TABLA 5.4 Variables indicativas de las ecuaciones de frontera de restricción para el problema de la Wyndor Glass Co.***

Restricción	Restricción en la forma aumentada	Ecuación de frontera de restricción	Variable indicativa
$x_1 \geq 0$	$x_1 \geq 0$	$x_1 = 0$	x_1
$x_2 \geq 0$	$x_2 \geq 0$	$x_2 = 0$	x_2
$x_1 \leq 4$	(1) $x_1 + x_3 = 4$	$x_1 = 4$	x_3
$2x_2 \leq 12$	(2) $2x_2 + x_4 = 12$	$2x_2 = 12$	x_4
$3x_1 + 2x_2 \leq 18$	(3) $3x_1 + 2x_2 + x_5 = 18$	$3x_1 + 2x_2 = 18$	x_5

* Variable indicativa $= 0 \Rightarrow$ se satisface la ecuación de frontera de restricción.
Variable indicativa $\neq 0 \Rightarrow$ se viola la ecuación de frontera de restricción.

TABLA 5.5 Soluciones BF para el problema de la Wyndor Glass Co.

Solución FEV	Ecuaciones de definición	Solución BF	Variables no básicas
(0, 0)	$x_1 = 0$ $x_2 = 0$	(0, 0, 4, 12, 18)	x_1 x_2
(0, 6)	$x_1 = 0$ $2x_2 = 12$	(0, 6, 4, 0, 6)	x_1 x_4
(2, 6)	$2x_2 = 12$ $3x_1 + 2x_2 = 18$	(2, 6, 2, 0, 0)	x_4 x_5
(4, 3)	$3x_1 + 2x_2 = 18$ $x_1 = 4$	(4, 3, 0, 6, 0)	x_5 x_3
(4, 0)	$x_1 = 4$ $x_2 = 0$	(4, 0, 0, 12, 6)	x_3 x_2

Cuando aumentan las soluciones FEV (vea la tabla 5.1) se obtienen las soluciones BF que se presentan en la tabla 5.5, en la cual se han colocado juntas las soluciones básicas factibles adyacentes, excepto el par formado por la primera y la última. Observe que, en todos los casos, las variables no básicas son necesariamente las variables indicativas de las ecuaciones de definición. Entonces, las soluciones BF adyacentes difieren en que tienen sólo una variable no básica distinta. También observe que cada solución BF es la solución simultánea del sistema de ecuaciones del problema en forma aumentada (vea la tabla 5.4) cuando las variables no básicas se igualan a cero.

De manera similar, las otras tres soluciones *no factibles* en los vértices (vea la tabla 5.2) conducen a las otras soluciones básicas *no factibles* que se muestran en la tabla 5.6.

Los otros dos conjuntos de variables no básicas, 1) x_1 y x_3 y 2) x_2 y x_4, no conducen a una solución básica porque al hacer cualquier par de variables igual a cero no se llega a tener una solución del sistema formado por las ecuaciones (1) a (3) de la tabla 5.4. Esta conclusión es paralela a la observación que se hizo antes respecto de que los conjuntos correspondientes de ecuaciones de frontera de restricción no conducen a una solución.

El *método símplex* comienza en una solución básica factible y se mueve en forma iterativa hacia una solución básica factible adyacente mejor, hasta que logra una solución óptima. ¿Cómo se alcanza la solución BF adyacente en cada iteración?

En el caso de la forma original del problema se debe recordar que se obtiene una solución factible en un vértice adyacente a partir de la solución actual cuando: 1) se elimina una restricción de frontera (ecuación de definición) del conjunto de n restricciones que definen la solución actual, 2) se hace un movimiento alejándose de la solución actual, en la dirección factible a lo largo de las $n - 1$ restricciones de frontera (una arista de la región factible) restantes y 3) el movimiento se detiene al encontrar la *primera* restricción (ecuación de definición) nueva.

De manera equivalente, en la terminología nueva, el método símplex llega a una solución BF adyacente a partir de la solución actual cuando: 1) se elimina una variable (la variable básica entrante) del conjunto de n variables no básicas que definen la solución actual, 2) se aleja de la so-

TABLA 5.6 Soluciones básicas no factibles para el problema de la Wyndor Glass Co.

Solución no factible en un vértice	Ecuaciones de definición	Solución básica no factible	Variables no básicas
(0, 9)	$x_1 = 0$ $3x_1 + 2x_2 = 18$	(0, 9, 4, −6, 0)	x_1 x_5
(4, 6)	$2x_2 = 12$ $x_1 = 4$	(4, 6, 0, 0, −6)	x_4 x_3
(6, 0)	$3x_1 + 2x_2 = 18$ $x_2 = 0$	(6, 0, −2, 12, 0)	x_5 x_2

TABLA 5.7 Secuencias de soluciones que se obtuvieron por el método símplex para el problema de la Wyndor Glass Co.

Iteración	Solución FEV	Ecuaciones de definición	Solution BF	Variables no básicas	Restricciones funcionales en la forma aumentada
0	(0, 0)	$x_1 = 0$ $x_2 = 0$	(0, 0, 4, 12, 18)	$x_1 = 0$ $x_2 = 0$	$x_1 + \mathbf{x_3} = 4$ $2x_2 + \mathbf{x_4} = 12$ $3x_1 + 2x_2 + \mathbf{x_5} = 18$
1	(0, 6)	$x_1 = 0$ $2x_2 = 12$	(0, 6, 4, 0, 6)	$x_1 = 0$ $x_4 = 0$	$x_1 + \mathbf{x_3} = 4$ $2\mathbf{x_2} + x_4 = 12$ $3x_1 + 2\mathbf{x_2} + \mathbf{x_5} = 18$
2	(2, 6)	$2x_2 = 12$ $3x_1 + 2x_2 = 18$	(2, 6, 2, 0, 0)	$x_4 = 0$ $x_5 = 0$	$\mathbf{x_1} + \mathbf{x_3} = 4$ $2\mathbf{x_2} + x_4 = 12$ $3\mathbf{x_1} + 2\mathbf{x_2} + x_5 = 18$

lución actual al *incrementar* el valor de esta variable a partir de cero (y al ajustar las otras variables básicas para que aún satisfagan el sistema de ecuaciones) al mismo tiempo que se mantienen las $n - 1$ variables no básicas restantes iguales cero y 3) se detiene cuando el valor de la *primera* variable básica (la variable básica saliente) llega a cero (a su restricción de frontera). Con cualquiera de estas dos interpretaciones, para elegir entre las n alternativas del paso 1 se selecciona aquella que proporciona la tasa más alta de mejoramiento del valor de Z (por cada unidad de incremento de la variable básica entrante) durante el paso 2.

En la tabla 5.7 se ilustra la cercana correspondencia entre estas interpretaciones geométrica y algebraica del método símplex. Con los resultados que se presentaron en las secciones 4.3 y 4.4, en la cuarta columna se resume la secuencia de soluciones BF del problema de la Wyndor Glass Co., y en la segunda columna se muestran las soluciones FEV correspondientes. En la tercera columna se puede observar que en cada iteración se elimina una frontera de restricción (ecuación de definición) y se incluye otra para obtener la nueva solución FEV. De manera similar, en la quinta columna se puede ver que cada iteración elimina una variable básica e incluye otra para obtener la nueva solución BF. Aún más, las variables no básicas que se quitan y se agregan son las variables indicativas de las ecuaciones de definición que se eliminan y se incluyen en la tercera columna. En la última columna se muestra el sistema de ecuaciones inicial [sin incluir la ecuación (0)] de la forma aumentada del problema, con las variables básicas actuales en **negritas**. En cada caso puede observarse cómo al hacer cero las variables no básicas y resolver el sistema de ecuaciones se obtiene la misma solución para (x_1, x_2) que con el par de ecuaciones de definición correspondiente en la tercera columna.

En la sección Worked Examples del sitio en internet de este libro se proporciona **otro ejemplo** en el que se desarrolla el tipo de información que se proporciona en la tabla 5.7 de un problema de minimización.

5.2 FORMA MATRICIAL DEL MÉTODO SÍMPLEX

En el capítulo 4 se describe el método símplex tanto en forma algebraica como tabular. Se puede obtener una visión más profunda de la teoría y del potencial del método símplex mediante el análisis de su forma *matricial*. Se comienza utilizando la notación matricial para representar problemas de programación lineal. (En el apéndice 4 se presenta un repaso del álgebra matricial.)

Para ayudar al lector a distinguir entre matrices, vectores y escalares, se usarán siempre letras **MAYÚSCULAS EN NEGRITAS** para representar matrices, letras **minúsculas en negritas** para representar vectores y letras *cursivas* normales en el caso de los escalares. También se usará el cero en negritas (**0**) para denotar el *vector nulo* (un vector cuyos elementos son todos iguales a cero) ya sea en forma de columna o de renglón (lo que debe ser claro por la forma del problema), mientras que el cero normal (0) seguirá representando al número cero.

Si se emplean matrices, nuestra forma estándar del modelo general de programación lineal establecido en la sección 3.2 se convierte en:

Maximizar	$Z = \mathbf{cx}$,	
sujeta a:		
$\mathbf{Ax} \leq \mathbf{b}$	y	$\mathbf{x} \geq \mathbf{0}$,

donde **c** es el vector renglón

$$\mathbf{c} = [c_1, c_2, \ldots, c_n],$$

x, **b** y **0** son vectores columna tales que

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \qquad \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

y **A** es la matriz

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Para obtener la *forma aumentada* del problema se introduce el vector columna de las variables de holgura

$$\mathbf{x}_s = \begin{bmatrix} x_{n+1} \\ x_{n+2} \\ \vdots \\ x_{n+m} \end{bmatrix}$$

de manera que las restricciones se convierten en

$$[\mathbf{A}, \mathbf{I}] \begin{bmatrix} \mathbf{x} \\ \mathbf{x}_s \end{bmatrix} = \mathbf{b} \quad \text{y} \quad \begin{bmatrix} \mathbf{x} \\ \mathbf{x}_s \end{bmatrix} \geq \mathbf{0},$$

donde **I** es la matriz identidad de orden $m \times m$ y el vector nulo **0** ahora tiene $n + m$ elementos. (Al final de la sección se hacen comentarios sobre cómo manejar problemas que no están en nuestra forma estándar.)

Obtención de una solución básica factible

En este punto es necesario recordar que el enfoque general del método símplex radica en obtener una secuencia de *soluciones BF mejoradas* hasta alcanzar la solución óptima. Una de las características clave de la forma matricial del método símplex revisado está relacionada con la forma en que obtiene cada nueva solución BF después de identificar sus variables básicas y no básicas. Dadas estas variables, la solución básica que resulta es la solución de las m ecuaciones

$$[\mathbf{A}, \mathbf{I}] \begin{bmatrix} \mathbf{x} \\ \mathbf{x}_s \end{bmatrix} = \mathbf{b},$$

en las que las *n variables no básicas* de entre los $n + m$ elementos de

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{x}_s \end{bmatrix}$$

se igualan a cero. Cuando se eliminan estas n variables al igualarlas a cero queda un conjunto de m ecuaciones con m incógnitas (las *variables básicas*). Este sistema de ecuaciones se puede denotar por

$$\mathbf{B}\mathbf{x}_B = \mathbf{b},$$

donde el **vector de variables básicas**

$$\mathbf{x}_B = \begin{bmatrix} x_{B1} \\ x_{B2} \\ \vdots \\ x_{Bm} \end{bmatrix}$$

se obtiene al eliminar las variables no básicas de

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{x}_s \end{bmatrix},$$

y la **matriz base**

$$\mathbf{B} = \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1m} \\ B_{21} & B_{22} & \dots & B_{2m} \\ \dots & \dots & \dots & \dots \\ B_{m1} & B_{m2} & \dots & B_{mm} \end{bmatrix}$$

se obtiene al eliminar las columnas correspondientes a los coeficientes de las variables no básicas de [**A**, **I**]. (Aún más, los elementos de $\mathbf{x}_B$ y, por lo tanto, las columnas de **B** pueden colocarse en orden diferente al ejecutar el método símplex.)

El método símplex introduce sólo variables básicas tales que **B** sea *no singular*, de manera que $\mathbf{B}^{-1}$ siempre existe. De esta forma, para resolver $\mathbf{B}\mathbf{x}_B = \mathbf{b}$, se premultiplican ambos lados por $\mathbf{B}^{-1}$:

$$\mathbf{B}^{-1}\mathbf{B}\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b}.$$

Como $\mathbf{B}^{-1}\mathbf{B} = \mathbf{I}$, la solución deseada para las variables básicas es

$$\boxed{\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b}.}$$

Sea $\mathbf{c}_B$ el vector cuyos elementos son los coeficientes de la función objetivo (incluye los ceros para las variables de holgura) que corresponden a los elementos de $\mathbf{x}_B$. El valor de la función objetivo de esta solución básica es, entonces,

$$\boxed{Z = \mathbf{c}_B\mathbf{x}_B = \mathbf{c}_B\mathbf{B}^{-1}\mathbf{b}.}$$

Ejemplo. Para ilustrar este método y obtener una solución básica factible, considere otra vez el problema de la Wyndor Glass Co., que se presentó en la sección 3.1, que se resolvió mediante el método símplex original en la tabla 4.8. En este caso,

$$\mathbf{c} = [3, 5], \quad [\mathbf{A}, \mathbf{I}] = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 1 & 0 \\ 3 & 2 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4 \\ 12 \\ 18 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{x}_s = \begin{bmatrix} x_3 \\ x_4 \\ x_5 \end{bmatrix}.$$

Con referencia a la tabla 4.8, la serie de soluciones básicas factibles que se obtiene mediante el método símplex es la siguiente:

Iteración 0

$$\mathbf{x}_B = \begin{bmatrix} x_3 \\ x_4 \\ x_5 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{B}^{-1} \text{ de manera que } \begin{bmatrix} x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 4 \\ 12 \\ 18 \end{bmatrix} = \begin{bmatrix} 4 \\ 12 \\ 18 \end{bmatrix},$$

$$\mathbf{c}_B = [0, 0, 0], \qquad \text{de manera que } Z = [0, 0, 0]\begin{bmatrix} 4 \\ 12 \\ 18 \end{bmatrix} = 0.$$

Iteración 1

$$\mathbf{x}_B = \begin{bmatrix} x_3 \\ x_2 \\ x_5 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 2 & 1 \end{bmatrix}, \qquad \mathbf{B}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & -1 & 1 \end{bmatrix},$$

de manera que

$$\begin{bmatrix} x_3 \\ x_2 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 12 \\ 18 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \\ 6 \end{bmatrix},$$

$$\mathbf{c}_B = [0, 5, 0], \qquad \text{entonces} \quad Z = [0, 5, 0] \begin{bmatrix} 4 \\ 6 \\ 6 \end{bmatrix} = 30.$$

Iteración 2

$$\mathbf{x}_B = \begin{bmatrix} x_3 \\ x_2 \\ x_1 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 2 & 3 \end{bmatrix}, \qquad \mathbf{B}^{-1} = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{bmatrix},$$

de manera que

$$\begin{bmatrix} x_3 \\ x_2 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 4 \\ 12 \\ 18 \end{bmatrix} = \begin{bmatrix} 2 \\ 6 \\ 2 \end{bmatrix},$$

$$\mathbf{c}_B = [0, 5, 3], \qquad \text{entonces} \quad Z = [0, 5, 3] \begin{bmatrix} 2 \\ 6 \\ 2 \end{bmatrix} = 36.$$

Forma matricial del conjunto de ecuaciones actual

La última consideración preliminar antes de resumir el método símplex revisado es mostrar la forma matricial del conjunto de ecuaciones que aparece en la tabla símplex en cualquier iteración del método símplex original.

En el caso del conjunto *original* de ecuaciones, la forma matricial es

$$\boxed{\begin{bmatrix} 1 & -\mathbf{c} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \mathbf{I} \end{bmatrix} \begin{bmatrix} Z \\ \mathbf{x} \\ \mathbf{x}_s \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{b} \end{bmatrix}.}$$

Este conjunto de ecuaciones se presenta también en la primera tabla símplex de la tabla 5.8.

Las operaciones algebraicas que se realizan por el método símplex (multiplicar una ecuación por una constante y sumar un múltiplo de una ecuación a otra) se expresan en forma matricial al premultiplicar ambos lados del conjunto original de ecuaciones por la matriz apropiada. Esta matriz tiene el mismo número de elementos que la matriz identidad, *excepto* que cada múltiplo de una operación algebraica se debe colocar en el lugar en que se necesita para que la multiplicación de matrices realice esta operación. Aun después de una serie de operaciones algebraicas a través

TABLA 5.8 Primera y última tabla símplex en forma matricial

Iteración	Variable básica	Ec.	Coeficiente de: Z	Variables originales	Variables de holgura	Lado derecho
0	Z $\mathbf{x}_B$	(0) (1, 2, . . . , m)	1 **0**	$-\mathbf{c}$ $\mathbf{A}$	$\mathbf{0}$ $\mathbf{I}$	0 $\mathbf{b}$
Cualquiera	Z $\mathbf{x}_B$	(0) (1, 2, . . . , m)	1 **0**	$\mathbf{c}_B\mathbf{B}^{-1}\mathbf{A}-\mathbf{c}$ $\mathbf{B}^{-1}\mathbf{A}$	$\mathbf{c}_B\mathbf{B}^{-1}$ $\mathbf{B}^{-1}$	$\mathbf{c}_B\mathbf{B}^{-1}\mathbf{b}$ $\mathbf{B}^{-1}\mathbf{b}$

de varias iteraciones se puede deducir cuál debe ser esta matriz (de manera simbólica) en cada paso, al usar lo que ya se sabe sobre el lado derecho del nuevo conjunto de ecuaciones. En particular, después de cualquier iteración, $\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b}$ y $Z = \mathbf{c}_B\mathbf{B}^{-1}\mathbf{b}$, por lo que el lado derecho de las ecuaciones se convierte en

$$\begin{bmatrix} Z \\ \mathbf{x}_B \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{c}_B\mathbf{B}^{-1} \\ \mathbf{0} & \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_B\mathbf{B}^{-1}\mathbf{b} \\ \mathbf{B}^{-1}\mathbf{b} \end{bmatrix}.$$

Debido a que se realiza la misma serie de operaciones algebraicas en *ambos* lados del conjunto original de ecuaciones, se usa esta misma matriz que premultiplica el lado derecho original para premultiplicar el lado izquierdo original. En consecuencia, como

$$\begin{bmatrix} 1 & \mathbf{c}_B\mathbf{B}^{-1} \\ \mathbf{0} & \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} 1 & -\mathbf{c} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{c}_B\mathbf{B}^{-1}\mathbf{A}-\mathbf{c} & \mathbf{c}_B\mathbf{B}^{-1} \\ \mathbf{0} & \mathbf{B}^{-1}\mathbf{A} & \mathbf{B}^{-1} \end{bmatrix},$$

la forma matricial que se busca para el *conjunto de ecuaciones después de cualquier iteración* es

$$\boxed{\begin{bmatrix} 1 & \mathbf{c}_B\mathbf{B}^{-1}\mathbf{A}-\mathbf{c} & \mathbf{c}_B\mathbf{B}^{-1} \\ \mathbf{0} & \mathbf{B}^{-1}\mathbf{A} & \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} Z \\ \mathbf{x} \\ \mathbf{x}_s \end{bmatrix} = \begin{bmatrix} \mathbf{c}_B\mathbf{B}^{-1}\mathbf{b} \\ \mathbf{B}^{-1}\mathbf{b} \end{bmatrix}.}$$

La segunda tabla símplex de la tabla 5.8 también muestra este mismo conjunto de ecuaciones.

Ejemplo. Para ilustrar esta forma matricial del conjunto actual de ecuaciones, considere el conjunto final de ecuaciones que se obtiene en la iteración 2 para el problema de la Wyndor Glass Co. Si se usan las $\mathbf{B}^{-1}$ y $\mathbf{c}_B$ dadas para la iteración 2 al final de la subsección anterior se tiene

$$\mathbf{B}^{-1}\mathbf{A} = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix},$$

$$\mathbf{c}_B\mathbf{B}^{-1} = [0, 5, 3] \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{bmatrix} = [0, \tfrac{3}{2}, 1],$$

$$\mathbf{c}_B\mathbf{B}^{-1}\mathbf{A} - \mathbf{c} = [0, 5, 3] \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} - [3, 5] = [0, 0].$$

Además, si se utilizan los valores de $\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b}$ y $Z = \mathbf{c}_B\mathbf{B}^{-1}\mathbf{b}$ que se calcularon al final de la subsección anterior, de estos resultados se obtiene el siguiente conjunto de ecuaciones:

$$\left[\begin{array}{c|cc|ccc} 1 & 0 & 0 & 0 & \frac{3}{2} & 1 \\ \hline 0 & 0 & 0 & 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 & -\frac{1}{3} & \frac{1}{3} \end{array}\right] \begin{bmatrix} Z \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 36 \\ 2 \\ 6 \\ 2 \end{bmatrix},$$

como se observa en la tabla símplex final de la tabla 4.8.

La forma matricial del conjunto de ecuaciones que queda después de cualquier iteración (como se muestra en el cuadro que está justo antes del ejemplo anterior) proporciona la clave para la ejecución de la forma matricial del método símplex. Las expresiones matriciales que muestran estas ecuaciones (o en la parte inferior de la tabla 5.8) proporcionan una forma directa de calcular todos los números que aparecerían en el conjunto de ecuaciones actual (para la forma algebraica del método símplex) o en la tabla símplex actual (para la forma tabular del método símplex). Las tres formas del método símplex realizan exactamente las mismas decisiones (variable básica entrante, variable básica saliente, etc.) paso a paso e iteración tras iteración. La única diferencia entre dichas formas radica en los métodos que se utilizan para calcular los números que se necesitan para realizar dichas decisiones. Como se resumirá en seguida, la forma matricial proporciona una forma conveniente y compacta de calcular dichos números sin tener que llevar a cabo una serie de sistemas de ecuaciones o de tablas símplex.

Resumen de la forma matricial del método símplex

1. *Inicialización*: Ingrese las variables de holgura, etc., para obtener las variables básicas iniciales como se describió en el capítulo 4. Lo anterior nos da $\mathbf{x}_B$, $\mathbf{c}_B$, $\mathbf{B}$ y $\mathbf{B}^{-1}$ (donde $\mathbf{B} = \mathbf{I} = \mathbf{B}^{-1}$ bajo nuestra suposición actual de que el problema que se pretende resolver se adapta a nuestra forma estándar). Después se procede a la prueba de optimalidad.

2. *Iteración*:

Paso 1. Determine la variable básica entrante: Remítase a los coeficientes de las variables no básicas de la ecuación (0) que se obtuvieron en la aplicación anterior de la prueba de optimalidad. Después (de la misma manera como se describió en la sección 4.4), seleccione la variable de *coeficiente negativo* que tenga el valor absoluto mayor como la variable básica entrante.

Paso 2. Determine la variable básica saliente: Utilice las expresiones matriciales $\mathbf{B}^{-1}\mathbf{A}$ (para los coeficientes de las variables orginales) y $\mathbf{B}^{-1}$ (para los coeficientes de las variables de holgura), para calcular los coeficientes de la variable básica entrante en cada una de las ecuaciones excepto la ecuación (0). Asimismo, utilice los cálculos anteriores de $\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b}$ (véase el paso 3) para identificar el lado derecho de dichas ecuaciones. Después (de la misma forma como se describió en la sección 4.4), utilice la *prueba del cociente mínimo* para seleccionar la variable básica saliente.

Paso 3. Determine la nueva solución BF: Actualice la matriz base $\mathbf{B}$, esto es, reemplace la columna de la variable básica saliente por la columna correspondiente en $[\mathbf{A}, \mathbf{I}]$ para la variable básica entrante. Asimismo, lleve a cabo los reemplazos correspondientes en $\mathbf{x}_B$ y $\mathbf{c}_B$. Después deduzca $\mathbf{B}^{-1}$ (como se ilustra en el apéndice 4) y fije el valor de $\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b}$.

3. *Prueba de optimalidad*: Use las expresiones matriciales, $\mathbf{c}_B\mathbf{B}^{-1}\mathbf{A} - \mathbf{c}$ (para los coeficientes de las variables originales) y $\mathbf{c}_B\mathbf{B}^{-1}$ (para los coeficientes de las variables de holgura), para calcular los coeficientes de las variables no básicas de la ecuación (0). La solución BF actual es óptima si y sólo si todos estos coeficientes son no negativos. Si la solución es óptima, deténgase. De otra forma, realice otra iteración a fin de obtener la solución BF siguiente.

Ejemplo. Ya hemos llevado a cabo algunos cálculos matriciales para el problema Wyndor Glass Co. dentro de esta sección. En seguida integraremos todas las partes con el fin de aplicar el método símplex en su totalidad en su forma matricial a este problema. Como punto de inicio recuerde que

$$\mathbf{c} = [3, 5], \quad [\mathbf{A, I}] = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 1 & 0 \\ 3 & 2 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4 \\ 12 \\ 18 \end{bmatrix}.$$

Inicialización

Las variables básicas iniciales son las variables de holgura, así que (como se pudo observar en la iteración 0 del primer ejemplo dentro de esta sección)

$$\mathbf{x}_B = \begin{bmatrix} x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 4 \\ 12 \\ 18 \end{bmatrix}, \quad \mathbf{c}_B = [0, 0, 0], \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \mathbf{B}^{-1}.$$

Prueba de optimalidad

Los coeficientes de las variables no básicas (x_1 y x_2) son

$$\mathbf{c}_B\mathbf{B}^{-1}\mathbf{A} - \mathbf{c} = [0, 0] - [3, 5] = [-3, -5]$$

por lo que dichos coeficientes negativos indican que la solución BF inicial ($\mathbf{x}_B = \mathbf{b}$) no es óptima.

Iteración 1

Puesto que –5 es mayor que –3 en valor absoluto, la variable básica entrante es x_2. Llevando a cabo sólo la porción relevante de la multiplicación matricial, los coeficientes de x_2 en todas las ecuaciones excepto en la (0) son

$$\mathbf{B}^{-1}\mathbf{A} = \begin{bmatrix} — & 0 \\ — & 2 \\ — & 2 \end{bmatrix}$$

y el lado derecho de dichas ecuaciones está dado por el valor de $\mathbf{x}_B$ que se muestra en la etapa de inicialización. Por lo tanto, la prueba del cociente mínimo muestra que la variable básica saliente es x_4 puesto que $12/2 < 18/2$. La iteración 1 del primer ejemplo en esta sección muestra los valores de $\mathbf{B}$, $\mathbf{x}_B$, $\mathbf{c}_B$ y $\mathbf{B}^{-1}$ actualizados, esto es:

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 2 & 1 \end{bmatrix}, \quad \mathbf{B}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & -1 & 1 \end{bmatrix}, \quad \mathbf{x}_B = \begin{bmatrix} x_3 \\ x_2 \\ x_5 \end{bmatrix} = \mathbf{B}^{-1}\mathbf{b} = \begin{bmatrix} 4 \\ 6 \\ 6 \end{bmatrix}, \mathbf{c}_B = [0, 5, 0],$$

por lo que x_2 reemplaza a x_4 en $\mathbf{x}_B$, para proporcionar un elemento de $\mathbf{c}_B$ de [3, 5, 0, 0, 0] y en proporcionar una columna de [**A**, **I**] en **B**.

Prueba de optimalidad

Las variables no básicas son ahora x_1 y x_4, y sus coeficientes en la ecuación (0) son

$$\text{Para } x_1: \qquad \mathbf{c}_B\mathbf{B}^{-1}\mathbf{A} - \mathbf{c} = [0, 5, 0] \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 3 & 2 \end{bmatrix} - [3, 5] = [-3, —, —]$$

$$\text{Para } x_4: \qquad \mathbf{c}_B\mathbf{B}^{-1} = [0, 5, 0] \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & -1 & 1 \end{bmatrix} = [—, 5/2, —]$$

Puesto que x_1 tiene un coeficiente negativo, la solución BF actual no es óptima, por lo que procedemos a realizar la siguiente iteración.

Iteración 2:
Puesto que x_1 es la variable no básica con un coeficiente negativo en la ecuación (0), ahora se convierte en la variable básica entrante. Sus coeficientes en las demás ecuaciones son

$$\mathbf{B}^{-1}\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & — \\ 0 & — \\ 3 & — \end{bmatrix}$$

Asimismo, utilizando la $\mathbf{x}_B$ obtenida al final de la iteración anterior, la prueba de cociente mínimo indica que x_5 es la variable básica de egreso, puesto que $6/3 < 4/1$. La iteración 2 del primer ejemplo de esta sección muestra los valores de $\mathbf{B}$, $\mathbf{B}^{-1}$, $\mathbf{x}_B$ y $\mathbf{c}_B$ actualizados que resultan,

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 2 & 3 \end{bmatrix}, \quad \mathbf{B}^{-1} = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}, \mathbf{x}_B = \begin{bmatrix} x_3 \\ x_2 \\ x_1 \end{bmatrix} = \mathbf{B}^{-1}\mathbf{b} = \begin{bmatrix} 2 \\ 6 \\ 2 \end{bmatrix}, \mathbf{c}_B = [0, 5, 3],$$

de tal forma que x_1 ha reemplazado a x_5 en $\mathbf{x}_B$, al proporcionar un elemento de $\mathbf{c}_B$ a partir de [3, 5, 0, 0, 0] y al proporcionar una columna de [**A**, **I**] en **B**.

Prueba de optimalidad
Las variables no básicas son ahora x_4 y x_5. Mediante el uso de los cálculos que se mostraron en el segundo ejemplo de esta sección, sus coeficientes en la ecuación (0) son 3/2 y 1, respectivamente. Puesto que ninguno de dichos coeficientes es negativo, la solución BF actual ($x_1 = 2, x_2 = 6, x_3 = 2, x_4 = 0, x_5 = 0$) es óptima por lo que aquí termina el proceso.

Observaciones finales

En el ejemplo anterior se muestra que la forma matricial del método símplex utiliza sólo algunas expresiones matriciales para llevar a cabo todos los cálculos necesarios. Dichas expresiones matriciales se resumen en la parte final de la tabla 5.8. Un hallazgo fundamental de esta tabla es que sólo se necesita conocer el $\mathbf{B}^{-1}$ y $\mathbf{c}_B\mathbf{B}^{-1}$ actuales, los cuales aparecen en la porción de las variables de holgura de la tabla símplex actual, con la finalidad de calcular todos los demás números de esta tabla en términos de los parámetros originales (**A**, **b** y **c**) del modelo que se pretende resolver. Cuando se trate de la tabla símplex *final*, este hallazgo es muy valioso, como se estudiará en la sección siguiente.

Una desventaja de la forma matricial del método símplex que se ha descrito en esta sección es que es necesario calcular $\mathbf{B}^{-1}$, la inversa de la matriz base actualizada, al final de cada iteración. A pesar de que existen rutinas para invertir matrices (no singulares) cuadradas pequeñas (y éstas pueden resolverse a mano en el caso de matrices de 2×2 o, inclusive en matrices de 3×3), el tiempo que se requiere para invertir matrices se incrementa muy rápido en función al tamaño de las matrices. Por fortuna, existe un procedimiento mucho más eficiente para actualizar $\mathbf{B}^{-1}$ de una iteración a la siguiente en lugar de invertir la matriz básica nueva desde el principio. Cuando este procedimiento se aplica a la forma matricial del método símplex, esta versión mejorada de la forma matricial se conoce convencionalmente con el nombre de **método símplex revisado**. Ésta es la versión del método símplex (junto con otras mejoras) que por lo general se utiliza en los paquetes de software comerciales de programación lineal. Se describirá el procedimiento de actualización de $\mathbf{B}^{-1}$ en la sección 5.4.

La sección de ejemplos del sitio en internet del libro proporciona **otro ejemplo** de la aplicación de la forma matricial del método símplex. Este ejemplo también incorpora un procedimiento eficiente para actualizar $\mathbf{B}^{-1}$ en cada iteración en lugar de invertir la matriz básica actualizada desde el principio, por lo que se aplica el método símplex revisado completo.

Por último, debemos recordarle que la descripción de la forma matricial del método símplex de esta sección supuso que el problema a resolver se adecua a *nuestra forma estándar* para el modelo de programación lineal general dado en la sección 3.2. Sin embargo, las modificaciones de las demás versiones del modelo son relativamente directas. La fase de inicialización se llevaría a cabo de la manera que se describió en la sección 4.6 tanto para la forma algebraica

como la tabular del método símplex. Cuando esta fase involucra la introducción de variables artificiales para obtener una solución BF inicial (y, por lo tanto, obtener una *matriz identidad como matriz base inicial*), estas variables se encuentran incluidas entre los m elementos de $\mathbf{x}_s$.

5.3 UNA IDEA FUNDAMENTAL

En esta sección se hará hincapié en una propiedad del método símplex (en cualquiera de sus formas) que el método símplex revisado puso de manifiesto en la sección anterior. Esta idea fundamental proporciona la clave tanto de la teoría de dualidad como del análisis de sensibilidad (capítulo 6), dos partes muy importantes de la programación lineal.

En primera instancia se describe esta idea cuando el problema que se está resolviendo se adecua a *nuestra forma estándar* para modelos de programación lineal (sección 3.2) y, después, se analiza cómo se adapta a otras formas. La idea se basa en la tabla 5.8 de la sección 5.2, como se describe a continuación.

Idea que se proporciona en la tabla 5.8: Mediante el uso de la notación matricial, la tabla 5.8 proporciona las filas de la tabla símplex *inicial* como $[-\mathbf{c}, \mathbf{0}, 0]$ para el renglón 0 y como $[\mathbf{A}, \mathbf{I}, \mathbf{b}]$ para el resto de los renglones. Después de cualquier iteración, los coeficientes de las variables de holgura en la tabla símplex actual se convierte en $\mathbf{c}_B\mathbf{B}^{-1}$ y en $\mathbf{B}^{-1}$ para los demás renglones, donde $\mathbf{B}$ es la matriz base actual. Si se examina el resto de la tabla símplex actual, la idea que surge es que dichos coeficientes de las variables de holgura inmediatamente ponen de manifiesto la forma en que *todos* los renglones de la tabla símplex actual se han obtenido a partir de los renglones en la tabla símplex *inicial*. En particular, después de cualquier iteración,

$$\text{Renglón } 0 = [-\mathbf{c}, \mathbf{0}, 0] + \mathbf{c}_B\mathbf{B}^{-1}[\mathbf{A}, \mathbf{I}, \mathbf{b}]$$
$$\text{Renglón } 1 \text{ a } m = \mathbf{B}^{-1}[\mathbf{A}, \mathbf{I}, \mathbf{b}]$$

Se describirán las aplicaciones de esta idea al final de esta sección. Dichas aplicaciones son particularmente importantes sólo cuando lidiamos con la tabla símplex *final*, una vez que se ha obtenido la solución óptima. Por lo tanto, se hará hincapié de aquí en adelante en el análisis de la “idea fundamental” sólo en términos de la solución óptima.

Con la finalidad de distinguir entre la notación matricial utilizada después de *cualquier* iteración ($\mathbf{B}^{-1}$, etc.) y la correspondiente notación justo después de la *última* iteración, presentamos la notación siguiente para el último caso.

Cuando $\mathbf{B}$ es la matriz base de la *solución óptima* calculada mediante el método símplex, sea

$\mathbf{S}^* = \mathbf{B}^{-1}$ = coeficientes de las variables *de holgura* de los renglones 1 a m
$\mathbf{A}^* = \mathbf{B}^{-1}\mathbf{A}$ = coeficientes de las variables *originales* de los renglones 1 a m
$\mathbf{y}^* = \mathbf{c}_B\mathbf{B}^{-1}$ = coeficientes de las variables *de holgura* del renglón 0
$\mathbf{z}^* = \mathbf{c}_B\mathbf{B}^{-1}\mathbf{A}$, de tal forma que $\mathbf{z}^* - \mathbf{c}$ = coeficientes de las variables *originales* del renglón 0
$Z^* = \mathbf{c}_B\mathbf{B}^{-1}\mathbf{b}$ = valor óptimo de la función objetivo
$\mathbf{b}^* = \mathbf{B}^{-1}\mathbf{b}$ = lados derechos óptimos de los renglones 1 a m

La mitad inferior de la tabla 5.9 muestra en qué parte de la tabla símplex final entra cada uno de estos símbolos. Para ilustrar la notación completa, la mitad superior de la tabla 5.9 incluye la tabla inicial del problema de la Wyndor Glass Co. y la mitad inferior incluye la tabla final de dicho problema.

Remítase de nuevo a este punto y suponga que le proporcionan a usted la tabla inicial, $\mathbf{t}$ y $\mathbf{T}$, y sólo $\mathbf{y}^*$ y $\mathbf{S}^*$ de la tabla final. ¿Cómo puede utilizarse únicamente esta información para calcular el resto de la tabla final? La respuesta se encuentra en la idea fundamental que se resume a continuación.

Idea fundamental

(1) $\mathbf{t}^* = \mathbf{t} + \mathbf{y}^*\mathbf{T} = [\mathbf{y}^*\mathbf{A} - \mathbf{c} \quad \mathbf{y}^* \mid \mathbf{y}^*\mathbf{b}]$.
(2) $\mathbf{T}^* = \mathbf{S}^*\mathbf{T} = [\mathbf{S}^*\mathbf{A} \quad \mathbf{S}^* \mid \mathbf{S}^*\mathbf{b}]$.

TABLA 5.9 Notificación general de las tablas símplex inicial y final en forma matricial para el problema de la Wyndor Glass Co.

Tabla inicial

Renglón 0: $\mathbf{t} = [-3, \ -5 \ \vdots \ 0, 0, 0 \quad 0] = [-\mathbf{c} \ \vdots \ \mathbf{0} \ \vdots \ 0]$.

Otros renglones: $\mathbf{T} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 4 \\ 0 & 2 & 0 & 1 & 0 & 12 \\ 3 & 2 & 0 & 0 & 1 & 18 \end{bmatrix} = [\mathbf{A} \ \vdots \ \mathbf{I} \quad \mathbf{b}]$.

Combinado: $\begin{bmatrix} \mathbf{t} \\ \mathbf{T} \end{bmatrix} = \begin{bmatrix} -\mathbf{c} & \mathbf{0} & 0 \\ \mathbf{A} & \mathbf{I} & \mathbf{b} \end{bmatrix}$.

Tabla final

Renglón 0: $\mathbf{t}^* = [0, 0 \ \vdots \ 0, \tfrac{3}{2}, 1 \ \vdots \ 36] = [\mathbf{z}^* - \mathbf{c} \ \vdots \ \mathbf{y}^* \ \vdots \ Z^*]$.

Otros renglones: $\mathbf{T}^* = \begin{bmatrix} 0 & 0 & 1 & \tfrac{1}{3} & -\tfrac{1}{3} & 2 \\ 0 & 1 & 0 & \tfrac{1}{2} & 0 & 6 \\ 1 & 0 & 0 & -\tfrac{1}{3} & \tfrac{1}{3} & 2 \end{bmatrix} = [\mathbf{A}^* \quad \mathbf{S}^* \quad \mathbf{b}^*]$.

Combinado: $\begin{bmatrix} \mathbf{t}^* \\ \mathbf{T}^* \end{bmatrix} = \begin{bmatrix} \mathbf{z}^* - \mathbf{c} & \mathbf{y}^* & Z^* \\ \mathbf{A}^* & \mathbf{S}^* & \mathbf{b}^* \end{bmatrix}$.

Por lo tanto, conociendo los parámetros del modelo en la tabla inicial (**c**, **A** y **b**) y *sólo* los coeficientes de las variables de holgura en la tabla final (**y*** y **S***), estas ecuaciones permiten calcular *todos* los demás valores de la tabla final.

Resumamos la lógica matemática que existe detrás de las dos ecuaciones de la idea fundamental. Para llegar a la ecuación (2) recuerde que la secuencia completa de operaciones algebraicas que realiza mediante el método símplex (excluyendo las que se encuentran en el renglón 0) es equivalente a premultiplicar **T** por alguna matriz, llamada **M**. Por lo tanto,

$$\mathbf{T}^* = \mathbf{MT},$$

sin embargo, ahora necesitamos identificar **M**. Reescribiendo las partes componentes de **T** y **T***, esta ecuación se transforma en

$$\begin{aligned} [\mathbf{A}^* \ \vdots \ \mathbf{S}^* \ \vdots \ \mathbf{b}^*] &= \mathbf{M}\,[\mathbf{A} \ \vdots \ \mathbf{I} \quad \mathbf{b}] \\ &= [\mathbf{MA} \ \vdots \ \mathbf{M} \ \vdots \ \mathbf{Mb}]. \end{aligned}$$

Debido a que el componente en la mitad (o en cualquier otra parte) de estas matrices iguales deben ser las mismas, se deduce que $\mathbf{M} = \mathbf{S}^*$, de tal forma que la ecuación (2) es válida.

La ecuación (1) se deduce de manera similar si se observa que toda la secuencia de operaciones algebraicas que involucren al renglón 0 se forma sumando alguna combinación lineal de los renglones de **T** a **t**, lo cual es equivalente a sumarle a **t** algún *vector* multiplicado por **T**. Representando este vector mediante la letra **v**, se tiene

$$\mathbf{t}^* = \mathbf{t} + \mathbf{vT},$$

pero aún es necesario que **v** sea identificado. Escribiendo de nuevo las partes componentes de **t** y **t*** obtenemos

$$\begin{aligned} [\mathbf{z}^* - \mathbf{c} \ \vdots \ \mathbf{y}^* \ \vdots \ Z^*] &= [-\mathbf{c} \ \vdots \ \mathbf{0} \ \vdots \ 0] + \mathbf{v}\,[\mathbf{A} \ \vdots \ \mathbf{I} \ \vdots \ \mathbf{b}] \\ &= [-\mathbf{c} + \mathbf{vA} \ \vdots \ \mathbf{v} \ \vdots \ \mathbf{vb}]. \end{aligned}$$

Igualando el componente del medio de estos vectores idénticos obtenemos $\mathbf{v} = \mathbf{y}^*$, lo cual valida la ecuación (1).

Adaptación de otras formas de modelos

Hasta ahora, la idea fundamental se ha descrito bajo el supuesto de que el modelo original se encuentra en nuestra forma estándar de la sección 3.2. No obstante, la lógica matemática anterior revela con exactitud qué ajustes se necesitan para otras formas del modelo original. La clave es la matriz identidad **I** de la tabla símplex inicial, la que se convierte en **S*** en la tabla símplex final. Si es necesario introducir algunas variables artificiales en la tabla símplex inicial para que sirvan de variables básicas iniciales, entonces **I** en esta tabla símplex está formada por el conjunto de columnas (ordenadas en forma apropiada) de *todas* las variables básicas iniciales (tanto de holgura como artificiales). (Las columnas de cualesquiera variables de exceso son irrelevantes.) Las *mismas* columnas de la tabla símplex final proporcionan **S*** de la ecuación $\mathbf{T}^* = \mathbf{S}^*\mathbf{T}$ y $\mathbf{y}^*$ de la ecuación $\mathbf{t}^* = \mathbf{t} + \mathbf{y}^*\mathbf{T}$. Si se introdujeran algunos valores de M en el renglón 0 preliminar como coeficientes de las variables artificiales, entonces la **t** de la ecuación $\mathbf{t}^* = \mathbf{t} + \mathbf{y}^*\mathbf{T}$ es el renglón 0 de la tabla símplex inicial después de eliminar en forma algebraica estos coeficientes distintos de cero de las variables básicas. (De otra manera, el renglón 0 preliminar se puede usar como **t**, pero deben restarse estas M del renglón 0 final para obtener $\mathbf{y}^*$.) (Vea el problema 5.3-9.)

Aplicaciones

La idea fundamental tiene una gran variedad de aplicaciones en programación lineal. Una de ellas involucra el método símplex revisado, que se basa principalmente en la forma matricial del método símplex presentado en la sección 5.2. Según se describió en la sección anterior (vea la tabla 5.8), este método utiliza $\mathbf{B}^{-1}$ y la tabla símplex inicial para calcular todos los números relevantes de la tabla símplex actual en *cada* iteración. Va aún más lejos que la idea fundamental al utilizar $\mathbf{B}^{-1}$ para calcular el mismo $\mathbf{y}^*$ como $\mathbf{y}^* = \mathbf{c}_B\mathbf{B}^{-1}$.

Otra aplicación incluye la interpretación de los *precios sombra* ($y_1^*, y_2^*, \ldots, y_m^*$) descritos en la sección 4.7. La idea fundamental revela que Z^* (el valor de Z para la solución óptima) es

$$Z^* = \mathbf{y}^*\mathbf{b} = \sum_{i=1}^{m} y_i^* b_i,$$

así, por ejemplo,

$$Z^* = 0b_1 + \frac{3}{2}b_2 + b_3$$

para el problema de la Wyndor Glass Co. Esta ecuación conduce de inmediato a la interpretación de las y_i^* dada en la sección 4.7.

Otro grupo de aplicaciones importantes en extremo incluye varias de las *tareas de posoptimalidad* (la técnica de reoptimización, el análisis de sensibilidad y la programación lineal paramétrica, descritas en la sección 4.7) que investigan el efecto que causan uno o más cambios en el modelo original. En particular, suponga que ya se aplicó el método símplex para obtener una solución óptima (y también $\mathbf{y}^*$ y $\mathbf{S}^*$) del modelo original y que estos cambios se hacen después. ¿Cuáles serían los cambios resultantes en la tabla símplex final si se aplicara en forma exacta la misma secuencia de operaciones algebraicas a la tabla inicial revisada? Como $\mathbf{y}^*$ y $\mathbf{S}^*$ no cambian, la idea fundamental revela de inmediato la respuesta.

Un tipo particularmente usual de análisis de posoptimalidad involucra investigar posibles cambios en **b**. A menudo, los elementos de **b** representan decisiones gerenciales acerca de la cantidad de diferentes recursos que se ponen a disposición de las actividades bajo consideración en el modelo de programación lineal. Por lo tanto, una vez que se ha obtenido la solución óptima mediante el método símplex, a menudo la gerencia desea conocer qué hubiera pasado si algunas de estas decisiones gerenciales acerca de la asignación de recursos se fueran a modificar de diferentes formas. Mediante el uso de las fórmulas,

$$\mathbf{x}_B = \mathbf{S}^*\mathbf{b}$$
$$\mathbf{Z}^* = \mathbf{y}^*\mathbf{b},$$

usted puede ver exactamente cómo varía la solución BF óptima (o si ésta se hace no factible debido a las variables negativas), así como la forma en que el valor óptimo de la función objetivo varía en función a **b**. Usted *no* tiene que volver a aplicar el método símplex una y otra vez para cada nueva **b**, ya que los coeficientes de las variables de holgura ¡lo dicen todo!

Por ejemplo, considere el cambio de $b_2 = 12$ a $b_2 = 13$ tal como se ilustra en la figura 4.8 en el problema de la Wyndor Glass Co. No es necesario resolver todo el problema para obtener la nueva solución óptima $(x_1, x_2) = (\frac{5}{3}, \frac{13}{2})$ porque la idea fundamental proporciona en seguida los valores de las variables básicas ($\mathbf{b}^*$) en la tabla símplex final:

$$\begin{bmatrix} x_3 \\ x_2 \\ x_1 \end{bmatrix} = \mathbf{b}^* = \mathbf{S}^*\mathbf{b} = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 4 \\ 13 \\ 18 \end{bmatrix} = \begin{bmatrix} \frac{7}{3} \\ \frac{13}{2} \\ \frac{5}{3} \end{bmatrix}.$$

Existe todavía una manera más sencilla de efectuar estos cálculos. Como el único cambio ocurre en la *segunda* componente de $\mathbf{b}$ ($\Delta b_2 = 1$), y este vector está premultiplicado nada más por la *segunda* columna de S^*, el *cambio* en $\mathbf{b}^*$ se puede calcular tan fácil como

$$\Delta\mathbf{b}^* = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{2} \\ -\frac{1}{3} \end{bmatrix} \Delta b_2 = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{2} \\ -\frac{1}{3} \end{bmatrix},$$

por lo que los valores originales de las variables básicas de la tabla símplex final ($x_3 = 2$, $x_2 = 6$, $x_1 = 2$) ahora se convierten en

$$\begin{bmatrix} x_3 \\ x_2 \\ x_1 \end{bmatrix} = \begin{bmatrix} 2 \\ 6 \\ 2 \end{bmatrix} + \begin{bmatrix} \frac{1}{3} \\ \frac{1}{2} \\ -\frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{7}{13} \\ \frac{13}{2} \\ \frac{5}{3} \end{bmatrix}.$$

(Si cualquiera de estos nuevos valores fuera *negativo* y, por lo tanto, no factible, se debería aplicar la técnica de reoptimización descrita en la sección 4.7; se iniciaría con esta tabla final revisada.) Al aplicar el *análisis incremental* a la ecuación anterior de Z^* también se obtiene

$$\Delta Z^* = \frac{3}{2}\Delta b_2 = \frac{3}{2}.$$

La idea fundamental también se puede aplicar a la investigación de otro tipo de cambios en el modelo original de una manera muy parecida; ésta es la parte crucial del procedimiento de análisis de sensibilidad que se describirá en la última parte del capítulo 6.

En el siguiente capítulo también se podrá ver que la idea fundamental tiene un papel clave en la teoría de dualidad para programación lineal.

5.4 EL MÉTODO SÍMPLEX REVISADO

El método símplex revisado se basa directamente en la forma matricial del método símplex que se presentó en la sección 5.2. Sin embargo, como se mencionó al final de dicha sección, la diferencia es que el método símplex revisado incorpora una mejora clave a la forma matricial. En lugar de que sea necesario invertir la nueva matriz base $\mathbf{B}$ después de cada iteración, lo cual es muy costoso desde el punto de vista computacional para matrices de gran tamaño, el método símplex revisado utiliza un procedimiento mucho más eficiente que simplemente actualiza $\mathbf{B}^{-1}$ de una iteración a la siguiente. Haremos hincapié en la descripción e ilustración de este procedimiento en esta sección.

Este procedimiento se basa en dos propiedades del método símplex. Una de ellas se describe en la *idea que se proporciona en la tabla 5.8* al comienzo de la sección 5.3. En particular, después de cualquier iteración, los coeficientes de las *variables de holgura* en todos los renglones excepto del 0 en la tabla símplex actual se hacen igual a $\mathbf{B}^{-1}$, donde $\mathbf{B}$ es la matriz base actual. Esta propiedad es válida siempre y cuando el problema que se trata de resolver se ajuste a *nuestra forma estándar* que se describió en la sección 3.2 de modelos de programación lineal. (Para formas no estándares donde sea necesario introducir variables artificiales, la única diferencia es que es el conjunto de columnas adecuadamente ordenadas que forme una matriz identidad $\mathbf{I}$ debajo del renglón 0 en la tabla símplex inicial, después proporciona $\mathbf{B}^{-1}$ en cualquier tabla subsecuente.)

La otra propiedad relevante del método símplex es que el paso 3 de una iteración cambia los números de la tabla símplex, incluyendo los números que generan $\mathbf{B}^{-1}$, sólo si se realizan las operaciones algebraicas elementales (tales como dividir una ecuación entre una constante o restar

un múltiplo de alguna ecuación de otra ecuación) que sean necesarias para recuperar la forma apropiada a partir de la eliminación gaussiana. Por lo tanto, todo lo que se necesita para actualizar $\mathbf{B}^{-1}$ de una iteración a la siguiente es obtener la nueva $\mathbf{B}^{-1}$ (representada por $\mathbf{B}^{-1}_{\text{nueva}}$) a partir de la $\mathbf{B}^{-1}$ vieja (representada por $\mathbf{B}^{-1}_{\text{vieja}}$) realizando las operaciones algebraicas usuales en $\mathbf{B}^{-1}_{\text{vieja}}$ que la forma alegabraica del método símplex llevaría a cabo en todo el sistema de ecuaciones [excepto en la ecuación (0)] para esta iteración. Por lo tanto, después de definir a la variable básica entrante y a la variable básica saliente de los pasos 1 y 2 de una iteración, el procedimiento consiste en aplicar el paso 3 de una iteración (como se describió en las secciones 4.3 y 4.4) a la porción $\mathbf{B}^{-1}$ de la tabla símplex actual o sistema de ecuaciones.

Para describir este procedimiento formalmente, sea

$x_k =$ variable básica entrante
$a'_{ik} =$ coeficiente de x_k en la ecuación (i) actual, para $i = 1,2, \ldots, m$ (identificada en el paso 2 de una iteración),
$r =$ número de la ecuación que contiene la variable básica saliente.

Recuerde que el nuevo conjunto de ecuaciones [excluyendo la ecuación (0)] puede obtenerse a partir del conjunto anterior restando a'_{ik}/a'_{rk} veces la ecuación (r) de la ecuación (i), para toda $i = 1, 2, \ldots, m$ excepto $i = r$ y después dividir la ecuación (r) por a'_{rk}. Por lo tanto, el elemento en el renglón i y la columna j de $\mathbf{B}^{-1}_{\text{nueva}}$ es

$$(\mathbf{B}^{-1}_{\text{nueva}})_{ij} = \begin{cases} (\mathbf{B}^{-1}_{\text{vieja}})_{ij} - \dfrac{a'_{ik}}{a'_{rk}}(\mathbf{B}^{-1}_{\text{vieja}})_{rj} & \text{si } i \neq r, \\ \dfrac{1}{a'_{rk}}(\mathbf{B}^{-1}_{\text{vieja}})_{rj} & \text{si } i = r. \end{cases}$$

Estas fórmulas pueden expresarse en notación matricial como

$$\mathbf{B}^{-1}_{\text{nueva}} = \mathbf{E}\mathbf{B}^{-1}_{\text{vieja}},$$

donde la matriz $\mathbf{E}$ es una matriz identidad excepto que si la columna r-ésima está reemplazada por el vector

$$\boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_m \end{bmatrix}, \qquad \text{donde} \qquad \eta_i = \begin{cases} -\dfrac{a'_{ik}}{a'_{rk}} & \text{si } i \neq r, \\ \dfrac{1}{a'_{rk}} & \text{si } i = r. \end{cases}$$

Así, $\mathbf{E} = [\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_{r-1}, \boldsymbol{\eta}, \mathbf{U}_{r+1}, \ldots, \mathbf{U}_m]$, donde los m elementos de cada vector columna $\mathbf{U}_i$ son 0 a excepción de un 1 en la i-ésima posición.

Ejemplo. Se ilustrará este procedimiento aplicándolo al problema de la compañía Wyndor Glass Co. Ya hemos aplicado la forma matricial del método símplex a este mismo problema en la sección 5.2, de tal forma que nos referiremos a los resultados obtenidos en cada iteración (la variable básica entrante, la variable básica saliente, etc.) para obtener la informacion necesaria para aplicar el procedimiento.

Iteración 1

Se pudo observar en la sección 5.2 que la $\mathbf{B}^{-1}$ inicial es igual a $\mathbf{I}$, la variable básica de ingreso es x_2 (de tal forma que $k = 2$), los coeficientes de x_2 en las ecuaciones 1, 2 y 3 son $a_{12} = 0$, $a_{22} = 2$ y $a_{32} = 2$, la variable básica saliente es x_4, y el número de la ecuación que contiene x_4 es $r = 2$. Para obtener la nueva $\mathbf{B}^{-1}$,

$$\boldsymbol{\eta} = \begin{bmatrix} -\dfrac{a_{12}}{a_{22}} \\ \dfrac{1}{a_{22}} \\ -\dfrac{a_{32}}{a_{22}} \end{bmatrix} = \begin{bmatrix} 0 \\ \dfrac{1}{2} \\ -1 \end{bmatrix},$$

de tal forma que

$$\mathbf{B}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

Iteración 2

Se pudo observar en la sección 5.2 que la variable básica entrante en esta iteración es x_1 (de tal forma que $k = 1$), los coeficientes de x_1 en las ecuaciones 1, 2 y 3 actuales son $a'_{11} = 1$, $a'_{21} = 0$ y $a'_{31} = 3$, la variable básica saliente es x_5 y el número de la ecuación que contiene a x_5 es $r = 3$. Estos resultados nos dan

$$\boldsymbol{\eta} = \begin{bmatrix} -\dfrac{a'_{11}}{a'_{31}} \\ -\dfrac{a'_{21}}{a'_{31}} \\ \dfrac{1}{a'_{31}} \end{bmatrix} = \begin{bmatrix} -\dfrac{1}{3} \\ 0 \\ \dfrac{1}{3} \end{bmatrix}$$

Por lo tanto, la $\mathbf{B}^{-1}$ nueva es

$$\mathbf{B}^{-1} = \begin{bmatrix} 1 & 0 & -\frac{1}{3} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

No son necesarias más iteraciones por ahora, por lo que este ejemplo ha finalizado.

Puesto que el método símplex revisado consiste en combinar este procedimiento para actualizar $\mathbf{B}^{-1}$ en cada iteración con el resto de la forma matricial del método símplex que se presentó en la sección 5.2, la combinación de este ejemplo con el de la sección 5.2 aplicando la forma matricial al mismo problema proporciona un ejemplo muy completo de cómo aplicar el método símplex revisado. Como se mencionó al final de la sección 5.2, la sección de Worked Examples del sitio en internet de este libro también proporciona **otro ejemplo** de cómo aplicar el método símplex revisado.

Concluimos esta sección con un resumen de las ventajas del método símplex revisado con respecto a la forma algebraica o tabular del método símplex. Una de ellas es que puede reducirse el número de cálculos aritméticos. Ello es especialmente cierto cuando la matriz **A** contiene un gran número de elementos igual a cero (lo cual por lo general es el caso en problemas de gran tamaño que se presentan en la práctica). La cantidad de información que debe almacenarse en cada iteración es menor y, a veces, mucho menor. El método símplex revisado también permite el control de los errores de redondeo que inevitablemente generan las computadoras. Dicho control puede ejecutarse si se obtiene de manera periódica al valor actual de $\mathbf{B}^{-1}$ invirtiendo directamente **B**. Además, algunos de los problemas del análisis de postoptimalidad que se analizaron en la sección 4.7 y al final de la 5.3 pueden manejarse de una forma más conveniente mediante el uso del método símplex revisado. Por todas estas razones, el método símplex revisado por lo general es la forma de método símplex que se prefiere para su ejecución por computadora.

5.5 CONCLUSIONES

Aunque el método símplex es un procedimiento algebraico está basado en algunos conceptos geométricos bastante sencillos. Estos conceptos permiten usar el algoritmo para examinar un número relativamente pequeño de soluciones básicas factibles antes de alcanzar e identificar la solución óptima.

En el capítulo 4 se describe cómo se utilizan las *operaciones algebraicas elementales* para ejecutar la *forma algebraica* del método símplex, y después el modo en que la *forma tabular* del méto-

do símplex utiliza las *operaciones elementales en los renglones* equivalentes de la misma manera. El estudio del método símplex en estas formas es una buena manera de iniciar el aprendizaje de los conceptos básicos. Sin embargo, estas formas del método símplex no proporcionan la manera más eficiente de ejecutarlo en computadora. Las *operaciones con matrices* son un camino más rápido para combinar y realizar operaciones algebraicas elementales u operaciones con renglones. Por lo tanto, al usar la *forma matricial* del método símplex, proporciona una manera eficaz de adaptar el método símplex para programarlo en una computadora. El *método símplex revisado* proporciona una mejora adicional para su implementación por computadora mediante la combinación de la forma matricial del método símplex con un procedimiento eficiente para actualizar la inversa de la matriz básica actual de iteración en iteración.

La tabla símplex final incluye la información completa para poderla reconstruir de manera algebraica a partir de la tabla símplex inicial. Esta idea fundamental tiene aplicaciones muy importantes, en especial en el análisis de posoptimalidad.

REFERENCIAS SELECCIONADAS

1. Dantzig, G. B. y M. N. Thapa, *Linear Programming 1: Introduction*, Springer, Nueva York, 1997.
2. Dantzig, G. B. y M. N. Thapa, *Linear Programming 2: Theory and Extensions*, Springer, Nueva York, 2003.
3. Luenberger, D. y Y. Ye: *Linear and Nolinear Programming*, 3a. ed. Springer, Nueva York, 2008.
4. Vanderbei, R. J., *Linear Programming*: *Foundations and Extensions*, 3a. ed., Springer, Nueva York, 2008.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN NUESTRO SITIO EN INTERNET (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos del capítulo 5

Ejemplo de demostración en el OR Tutor:

Idea fundamental

Rutinas interactivas en el IOR Tutorial:

Método gráfico interactivo
Introducción o revisión de un modelo general de programación lineal
Preparación para el método símplex, sólo interactivo
Solución interactiva mediante el método símplex

Rutinas automáticas en el IOR Tutorial:

Solución automática mediante el método símplex
Método gráfico y análisis de sensibilidad

Archivos (capítulo 3) para resolver el ejemplo Wyndor:

Archivos de Excel
Archivo LINGO/LINDO
Archivo MPL/CPLEX

Glosario del capítulo 5

Vea el apéndice 1 para la documentación del software.

PROBLEMAS

Los símbolos a la izquierda de algunos problemas (o de sus incisos) significan lo siguiente:

D: El ejemplo de demostración que se presentó antes puede ser útil.
I: Puede verificar algunos de sus trabajos por medio de los procedimientos listados anteriormente.

Un asterisco en el número del problema indica que al final del libro se da al menos una respuesta parcial.

5.1-1.* Considere el siguiente problema.

Maximizar $Z = 3x_1 + 2x_2$,

sujeta a

$$2x_1 + x_2 \leq 6$$
$$x_1 + 2x_2 \leq 6$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

I ***a***) Resuelva este problema en forma gráfica. Identifique las soluciones FEV en la gráfica encerrándolas en un círculo.

b) Identifique todos los conjuntos de dos ecuaciones de definición del problema. En cada uno obtenga (si existe) la solución correspondiente en el vértice y clasifíquela como solución FEV o como no factible.

c) Introduzca las variables de holgura con objeto de expresar las ecuaciones funcionales en la forma aumentada. Utilice estas variables de holgura para identificar la solución básica que corresponde a cada solución en un vértice, que haya encontrado en el inciso *b*).

d) Haga lo siguiente en *cada* conjunto de dos ecuaciones de definición del inciso *b*): identifique las variables indicativas de cada ecuación de definición; escriba los conjuntos de ecuaciones del inciso *c*) *después* de eliminar estas dos variables indicativas (no básicas); luego use el último conjunto de ecuaciones para obtener las dos variables restantes (variables básicas); compare la solución básica que resulta con la solución básica correspondiente que obtuvo en el inciso *c*).

e) Sin ejecutar el método símplex, utilice su interpretación geométrica (y la función objetivo) para identificar la trayectoria (secuencia de soluciones FEV) que seguiría para llegar a la solución óptima. En cada solución FEV identifique la decisión tomada para la siguiente iteración: *i*) cuál ecuación de definición se elimina y cuál se agrega; *ii*) cuál variable indicativa se elimina (variable básica entrante) y cuál se introduce (variable básica que sale).

5.1-2. Repita el problema 5.1-1 en el modelo del problema 3.1-6.

5.1-3. Considere el siguiente problema.

Maximizar $Z = 5x_1 + 8x_2$,

sujeta a

$$4x_1 + 2x_2 \leq 80$$
$$-3x_1 + x_2 \leq 4$$
$$-x_1 + 2x_2 \leq 20$$
$$4x_1 - x_2 \leq 40$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

I ***a***) Resuelva este problema en forma gráfica. Identifique las soluciones FEV y señálelas en la gráfica.

b) Desarrolle una tabla que muestre cada solución FEV y las ecuaciones de definición correspondientes, la solución BF y las variables no básicas. Calcule Z de cada solución y utilice sólo esta información para identificar la solución óptima.

c) Desarrolle la tabla correspondiente de las soluciones no factibles en los vértices, etc. Además, identifique los conjuntos de ecuaciones de definición y las variables no básicas que no conducen a una solución.

5.1-4. Considere el siguiente problema.

Maximizar $Z = 2x_1 - x_2 + x_3$,

sujeta a

$$3x_1 + x_2 + x_3 \leq 60$$
$$x_1 - x_2 + 2x_3 \leq 10$$
$$x_1 + x_2 - x_3 \leq 20$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

Después de introducir las variables de holgura y de realizar una iteración completa del método símplex se obtiene la siguiente tabla símplex.

Iteración	Variable básica	Ec.	Coeficiente de:							Lado derecho
			Z	x_1	x_2	x_3	x_4	x_5	x_6	
1	Z	(0)	1	0	−1	3	0	2	0	20
	x_4	(1)	0	0	4	−5	1	−3	0	30
	x_1	(2)	0	1	−1	2	0	1	0	10
	x_6	(3)	0	0	2	−3	0	−1	1	10

a) Identifique la solución FEV que se obtuvo mediante la iteración 1.

b) Identifique las ecuaciones de frontera de restricción que definen la solución FEV.

5.1-5. Considere el problema de programación lineal de tres variables que se muestra en la figura 5.2.

a) Construya una tabla similar a la 5.1 que muestre el conjunto de ecuaciones de definición de cada solución FEV.

b) ¿Cuáles son las ecuaciones de definición de la solución no factible en el vértice (6, 0, 5)?

c) Identifique uno de los sistemas de tres ecuaciones de frontera de restricción que no conduzca a una solución FEV ni a una solución no factible en un vértice. Explique por qué ocurre esto en ese sistema.

5.1-6. Considere el siguiente problema.

Minimizar $Z = 8x_1 + 5x_2$,

sujeta a

$$-3x_1 + 2x_2 \leq 30$$
$$2x_1 + x_2 \geq 50$$
$$x_1 + x_2 \geq 30$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

a) Identifique los 10 conjuntos de ecuaciones de definición del problema. Para cada uno, obtenga (si existe) la solución correspondiente en el vértice y clasifíquela como solución FEV o como solución no factible en un vértice.
b) En cada solución en un vértice, obtenga la solución básica correspondiente y su conjunto de variables no básicas.

5.1-7. Reconsidere el modelo del problema 3.1-5.
a) Identifique los 15 conjuntos de ecuaciones de definición de este problema. En cada uno, obtenga (si existe) la solución correspondiente en un vértice y clasifíquela como solución FEV o como solución no factible en un vértice.
b) En cada solución en un vértice obtenga la solución básica correspondiente y su conjunto de variables no básicas.

5.1-8. Cada una de las siguientes afirmaciones es cierta en casi todas las circunstancias, pero no siempre. En cada caso, indique cuándo no lo es y por qué.
a) La mejor solución FEV es una solución óptima.
b) Una solución óptima es una solución FEV.
c) Una solución FEV es la única solución óptima si ninguna de las soluciones FEV adyacentes es mejor (según la medida del valor de la función objetivo).

5.1-9. Considere la forma original (antes de aumentar) de un problema de programación lineal con *n* variables de decisión (cada una con una restricción de no negatividad) y *m* restricciones funcionales. Diga si las siguientes afirmaciones son falsas o verdaderas y después justifique su respuesta con referencias específicas al material del capítulo (incluya cita de la página).
a) Si una solución factible es óptima, debe ser una solución FEV.
b) El número de soluciones FEV es al menos

$$\frac{(m+n)!}{m!n!}.$$

c) Si una solución FEV tiene soluciones adyacentes que son mejores (de acuerdo con el valor de *Z*), entonces una de estas soluciones FEV adyacentes debe ser óptima.

5.1-10. Diga si las siguientes afirmaciones sobre problemas de programación lineal son falsas o verdaderas y después justifique su respuesta.
a) Si una solución factible es óptima pero no FEV, existe un número infinito de soluciones óptimas.
b) Si el valor de la función objetivo es igual en dos puntos factibles diferentes $\mathbf{x}^*$ y $\mathbf{x}^{**}$, entonces todos los puntos sobre el segmento de recta que conecta a $\mathbf{x}^*$ y $\mathbf{x}^{**}$ son factibles y *Z* tiene el mismo valor en todos ellos.
c) Si el problema tiene *n* variables (antes de aumentar), entonces la solución simultánea de cualquier conjunto de *n* ecuaciones de frontera de restricción es una solución FEV.

5.1-11. Considere la forma aumentada de los problemas de programación lineal que tienen soluciones factibles y una región factible acotada. Diga si cada una de las afirmaciones siguientes es falsa o verdadera y después justifique su respuesta con una referencia a afirmaciones específicas del capítulo (incluya cita de la página).
a) Debe haber al menos una solución óptima.
b) Una solución óptima debe ser una solución BF.
c) El número de soluciones BF es finito.

5.1-12.* Reconsidere el modelo del problema 4.6-9. Ahora se sabe que las variables básicas de la solución óptima son x_2 y x_3. Use esta información para identificar un sistema de tres ecuaciones de frontera de restricción cuya solución simultánea sea esta solución óptima. Resuelva este sistema de ecuaciones para obtener esa solución.

5.1-13. Reconsidere el problema 4.3-6. Ahora use la información dada y la teoría del método símplex para identificar un sistema de tres ecuaciones de frontera de restricción (en x_1, x_2, x_3) cuya solución simultánea sea la solución óptima, sin aplicar el método símplex. Resuelva el sistema de ecuaciones para encontrar la solución óptima.

5.1-14. Considere el siguiente problema.

$$\text{Maximizar} \qquad Z = 2x_1 + 2x_2 + 3x_3,$$

sujeta a

$$\begin{aligned} 2x_1 + x_2 + 2x_3 &\leq 4 \\ x_1 + x_2 + x_3 &\leq 3 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0.$$

Sean x_4 y x_5 las variables de holgura de las restricciones funcionales respectivas. Comience con estas dos variables como las variables básicas de la solución BF inicial, y se obtiene la información de que el método símplex procede de la siguiente manera para obtener la solución óptima en dos iteraciones: 1) en la iteración 1, la variable básica entrante es x_3 y la variable básica saliente es x_4; 2) en la iteración 2, la variable básica entrante es x_2 y la que sale es x_5.
a) Desarrolle un dibujo de tres dimensiones de la región factible del problema y muestre la trayectoria que sigue el método símplex.
b) Dé una interpretación geométrica de las razones por las cuales el método símplex sigue esta trayectoria.
c) En cada una de las dos aristas de la región factible por las que se desplaza el método símplex, dé la ecuación de las dos fronteras de restricción sobre las que está y después la ecuación de la frontera de restricción adicional en cada punto extremo.
d) Identifique el conjunto de ecuaciones de definición de las tres soluciones FEV (incluso la inicial) que obtuvo por el método símplex. Use las ecuaciones de definición para obtener estas soluciones.
e) En cada solución FEV que obtuvo en el inciso *d*), proporcione la solución BF correspondiente y su conjunto de variables no básicas. Explique en qué forma estas variables no básicas identifican las ecuaciones de definición del inciso *d*).

5.1-15. Considere el siguiente problema.

$$\text{Maximizar} \qquad Z = 3x_1 + 4x_2 + 2x_3,$$

sujeta a

$$\begin{aligned} x_1 + x_2 + x_3 &\leq 20 \\ x_1 + 2x_2 + x_3 &\leq 30 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0.$$

Sean x_4 y x_5 las variables de holgura de las restricciones funcionales respectivas. Si se inicia con estas dos variables como variables básicas de la solución BF inicial, se recibe la información de que el

método símplex procede de la siguiente manera para obtener la solución óptima en dos iteraciones: 1) en la iteración 1, la variable básica entrante es x_2 y la variable básica que sale es x_5; 2) en la iteración 2, la variable básica entrante es x_1 y la que sale es x_4.

Siga las mismas instrucciones del problema 5.1-14 para este caso.

5.1-16. Después de observar la figura 5.2, explique por qué la propiedad 1*b* de las soluciones FEV se cumple en este problema si tiene la siguiente función objetivo.

a) Maximizar $Z = x_3$.
b) Maximizar $Z = -x_1 + 2x_3$.

5.1-17. Considere el problema de programación lineal de tres variables que se muestra en la figura 5.2.

a) Explique en términos geométricos por qué el conjunto de soluciones que satisfacen cualquier restricción individual es convexo según se define en el apéndice 2.
b) Utilice la conclusión del inciso *a*) para explicar por qué la región factible completa (el conjunto de soluciones que satisfacen de manera simultánea todas las restricciones) es un conjunto convexo.

5.1-18. Suponga que el problema de programación lineal de tres variables que se presenta en la figura 5.2 tiene la función objetivo

$$\text{Maximizar} \qquad Z = 3x_1 + 4x_2 + 3x_3.$$

Sin emplear el álgebra del método símplex, aplique sólo el razonamiento geométrico (incluso la elección de la arista que proporcione la tasa máxima de incremento de *Z*) para determinar y explicar la trayectoria que seguiría en la figura 5.2 desde el origen hasta la solución óptima.

5.1-19. Considere el problema de programación lineal de tres variables que se muestra en la figura 5.2.

a) Construya una tabla similar a la 5.4 que muestre las variables indicativas de cada ecuación de frontera de restricción y cada restricción original.
b) Para la solución FEV (2, 4, 3) y sus tres soluciones FEV adyacentes (4, 2, 4), (0, 4, 2) y (2, 4, 0) construya una tabla similar a la 5.5 que muestre las ecuaciones de definición correspondientes, las soluciones BF y las variables no básicas.
c) Utilice los conjuntos de ecuaciones de definición del inciso *b*) para demostrar que (4, 2, 4), (0, 4, 2) y (2, 4, 0) son en realidad adyacentes a (2, 4, 3), pero que ninguna de estas tres soluciones FEV son adyacentes entre sí. Después utilice los conjuntos de variables no básicas del inciso *b*) para demostrar lo mismo.

5.1-20. La fórmula de la recta que pasa por (2, 4, 3) y (4, 2, 4) de la figura 5.2 se puede escribir como

$$(2, 4, 3) + \alpha[(4, 2, 4) - (2, 4, 3)] = (2, 4, 3) + \alpha(2, -2, 1),$$

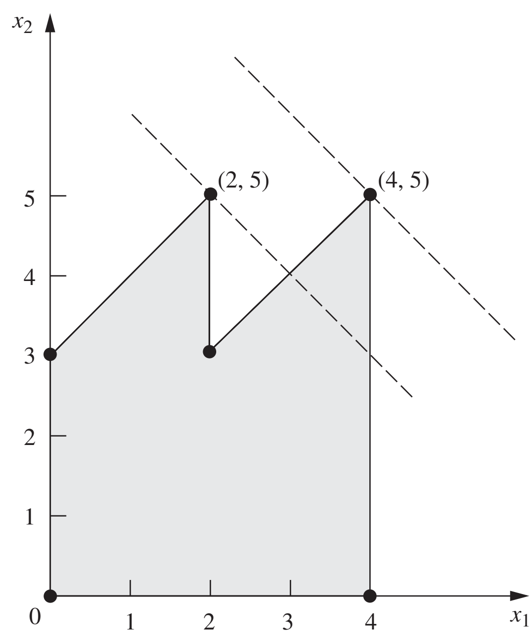
donde $0 \leq \alpha \leq 1$ para el segmento de recta que une estos dos puntos. Después de aumentar las restricciones respectivas con las variables de holgura x_4, x_5, x_6, x_7 esta fórmula se convierte en

$$(2, 4, 3, 2, 0, 0, 0) + \alpha(2, -2, 1, -2, 2, 0, 0).$$

Utilice esta fórmula directamente para contestar lo siguiente y relacionar el álgebra y la geometría del método símplex al ir de una iteración a otra, desde el punto (2, 4, 3) hasta (4, 2, 4). (Se tiene información de que se mueve a lo largo de este segmento de recta.)

a) ¿Cuál es la variable básica entrante?
b) ¿Cuál es la variable básica que sale?
c) ¿Cuál es la nueva solución BF?

5.1-21. Considere un problema de programación matemática con dos variables cuya región factible se muestra en la gráfica; los seis puntos corresponden a las soluciones FEV. El problema tiene una función objetivo lineal y las dos rectas punteadas indican rectas de función objetivo que pasan por la solución óptima (4, 5) y por la segunda mejor solución FEV (2, 5). Observe que la solución no óptima (2, 5) es mejor que las dos soluciones FEV adyacentes, lo cual viola la propiedad 3 de programación lineal que se explicó en la sección 5.1 para soluciones FEV. Construya la región factible que resultaría si los seis segmentos de la frontera fueran restricciones de frontera para las restricciones de programación lineal, con el fin de demostrar que este problema *no puede ser* un problema de programación lineal.



5.2-1. Considere el siguiente problema.

$$\text{Maximizar} \qquad Z = 8x_1 + 4x_2 + 6x_3 + 3x_4 + 9x_5,$$

sujeta a

$$\begin{aligned} x_1 + 2x_2 + 3x_3 + 3x_4 &\leq 180 \qquad \text{(recurso 1)} \\ 4x_1 + 3x_2 + 2x_3 + x_4 + x_5 &\leq 270 \qquad \text{(recurso 2)} \\ x_1 + 3x_2 + x_4 + 3x_5 &\leq 180 \qquad \text{(recurso 3)} \end{aligned}$$

y

$$x_j \geq 0, \qquad j = 1, \ldots, 5.$$

Se sabe que las variables básicas de la solución óptima son x_3, x_1 y x_5 y que

$$\begin{bmatrix} 3 & 1 & 0 \\ 2 & 4 & 1 \\ 0 & 1 & 3 \end{bmatrix}^{-1} = \frac{1}{27}\begin{bmatrix} 11 & -3 & 1 \\ -6 & 9 & -3 \\ 2 & -3 & 10 \end{bmatrix}.$$

a) Con esta información identifique la solución óptima.
b) Use esta información para identificar los precios sombra de los tres recursos.

I **5.2-2.*** Aplique la forma matricial del método símplex para resolver el siguiente problema.

Maximizar $Z = 5x_1 + 8x_2 + 7x_3 + 4x_4 + 6x_5$,

sujeta a

$$2x_1 + 3x_2 + 3x_3 + 2x_4 + 2x_5 \leq 20$$
$$3x_1 + 5x_2 + 4x_3 + 2x_4 + 4x_5 \leq 30$$

y

$$x_j \geq 0, \quad j = 1, 2, 3, 4, 5.$$

5.2-3. Reconsidere el problema 5.1-1. Para la sucesión de soluciones FEV que se identificó en el inciso *e*), construya la matriz base **B** de cada solución BF correspondiente. En cada caso, invierta **B** en forma manual y utilice esta $\mathbf{B}^{-1}$ para calcular la solución actual; después realice la siguiente iteración (o demuestre que la actual es óptima).

I **5.2-4.** Aplique la forma matricial del método símplex para resolver el modelo que se presentó en el problema 4.1-5.

I **5.2-5.** Aplique la forma matricial del método símplex para resolver el modelo que se presentó en el problema 4.7-6.

D **5.3-1.*** Considere el siguiente problema.

Maximizar $Z = x_1 - x_2 + 2x_3$,

sujeta a

$$2x_1 - 2x_2 + 3x_3 \leq 5$$
$$x_1 + x_2 - x_3 \leq 3$$
$$x_1 - x_2 + x_3 \leq 2$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

Sean x_4, x_5 y x_6 las variables de holgura de las restricciones respectivas. Después de aplicar el método símplex, una parte de la tabla símplex final es como se muestra a continuación:

Variable básica	Ec.	Coeficiente de:							Lado derecho
		Z	x_1	x_2	x_3	x_4	x_5	x_6	
Z	(0)	1				1	1	0	
x_2	(1)	0				1	3	0	
x_6	(2)	0				0	1	1	
x_3	(3)	0				1	2	0	

a) Utilice la idea fundamental presentada en la sección 5.3 para identificar los números que faltan en esta tabla símplex final. Muestre sus cálculos.

b) Identifique las ecuaciones de definición de la solución FEV correspondiente a la solución BF óptima de la tabla símplex final.

D **5.3-2.** Considere el siguiente problema.

Maximizar $Z = 4x_1 + 3x_2 + x_3 + 2x_4$,

sujeta a

$$4x_1 + 2x_2 + x_3 + x_4 \leq 5$$
$$3x_1 + x_2 + 2x_3 + x_4 \leq 4$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0.$$

Sean x_5 y x_6 las variables de holgura de las restricciones respectivas. Después de aplicar el método símplex, una parte de la tabla final es como se muestra a continuación:

Variable básica	Ec.	Coeficiente de:							Lado derecho
		Z	x_1	x_2	x_3	x_4	x_5	x_6	
Z	(0)	1					1	1	
x_2	(1)	0					1	−1	
x_4	(2)	0					−1	2	

a) Utilice la idea fundamental que se presentó en la sección 5.3 para identificar los números que faltan en esta tabla. Muestre sus cálculos.

b) Identifique las ecuaciones de definición de la solución FEV que corresponde a la solución BF óptima de la tabla símplex final.

D **5.3-3.** Considere el siguiente problema.

Maximizar $Z = 6x_1 + x_2 + 2x_3$,

sujeta a

$$2x_1 + 2x_2 + \frac{1}{2}x_3 \leq 2$$
$$-4x_1 - 2x_2 - \frac{3}{2}x_3 \leq 3$$
$$x_1 + 2x_2 + \frac{1}{2}x_3 \leq 1$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

Sean x_4, x_5 y x_6 las variables de holgura de las restricciones respectivas. Después de aplicar el método símplex, una parte de la tabla símplex final es como se muestra a continuación:

Variable básica	Ec.	Coeficiente de:							Lado derecho
		Z	x_1	x_2	x_3	x_4	x_5	x_6	
Z	(0)	1				2	0	2	
x_5	(1)	0				1	1	2	
x_3	(2)	0				−2	0	4	
x_1	(3)	0				1	0	−1	

Utilice la idea fundamental que se presentó en la sección 5.3 para identificar los números que faltan en esta tabla símplex final. Muestre sus cálculos.

D **5.3-4.** Considere el siguiente problema.

Maximizar $Z = 20x_1 + 6x_2 + 8x_3$,

sujeta a

$$\begin{aligned} 8x_1 + 2x_2 + 3x_3 &\leq 200 \\ 4x_1 + 3x_2 \qquad &\leq 100 \\ 2x_1 \qquad + x_3 &\leq 50 \\ x_3 &\leq 20 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0.$$

Sean x_4, x_5, x_6 y x_7 las variables de holgura de las restricciones respectivas. Después de aplicar el método símplex, una parte de la tabla símplex final es como se muestra a continuación:

Variable básica	Ec.	Coeficiente de:								Lado derecho
		Z	x_1	x_2	x_3	x_4	x_5	x_6	x_7	
Z	(0)	1				$\frac{9}{4}$	$\frac{1}{2}$	0	0	
x_1	(1)	0				$\frac{3}{16}$	$-\frac{1}{8}$	0	0	
x_2	(2)	0				$-\frac{1}{4}$	$\frac{1}{2}$	0	0	
x_6	(3)	0				$-\frac{3}{8}$	$\frac{1}{4}$	1	0	
x_7	(4)	0				0	0	0	1	

a) Utilice la idea fundamental que se presentó en la sección 5.3 para identificar los números que faltan en esta tabla símplex final. Muestre sus cálculos.

b) Indique cuáles de estos números se generarían con la forma matricial del método símplex para poder realizar la siguiente iteración.

c) Identifique las ecuaciones de definición de la solución FEV correspondiente a la solución BF de la tabla símplex actual.

D **5.3-5.** Considere el siguiente problema.

$$\text{Maximizar} \qquad Z = c_1x_1 + c_2x_2 + c_3x_3,$$

sujeta a

$$\begin{aligned} x_1 + 2x_2 + x_3 &\leq b \\ 2x_1 + x_2 + 3x_3 &\leq 2b \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0.$$

Observe que no se asignaron valores a los coeficientes de la función objetivo (c_1, c_2, c_3), y que la única especificación para los valores del lado derecho de las restricciones funcionales es que el segundo ($2b$) es el doble del primero (b).

Ahora suponga que su jefe decidió insertar su mejor estimación de los valores de c_1, c_2, c_3 y b sin informarle y después corrió el método símplex. El resultado es la tabla símplex final resultante que se muestra a continuación (donde x_4 y x_5 son las variables de holgura de las restricciones funcionales respectivas), pero no puede leer el valor de Z^*.

Variable básica	Ec.	Coeficiente de:						Lado derecho
		Z	x_1	x_2	x_3	x_4	x_5	
Z	(0)	1	$\frac{7}{10}$	0	0	$\frac{3}{5}$	$\frac{4}{5}$	Z^*
x_2	(1)	0	$\frac{1}{5}$	1	0	$\frac{3}{5}$	$-\frac{1}{5}$	1
x_3	(2)	0	$\frac{3}{5}$	0	1	$-\frac{1}{5}$	$\frac{2}{5}$	3

a) Utilice la idea fundamental que se presentó en la sección 5.3 para identificar el valor de (c_1, c_2, c_3) que se usó.

b) Utilice la idea fundamental que se presentó en la sección 5.3 para identificar el valor de b que se usó.

c) Calcule el valor de Z^* de dos maneras; en una de ellas utilice los resultados del inciso *a*) y en la otra los resultados del inciso *b*). Muestre los métodos para encontrar Z^*.

5.3-6. La siguiente expresión se obtuvo en la iteración 2 del ejemplo de la sección 5.3:

$$\text{Renglón final } 0 = [-3, \quad -5 \mid 0, \quad 0, \quad 0 \mid 0] + [0, \quad \tfrac{3}{2}, \quad 1] \left[\begin{array}{cc|ccc|c} 1 & 0 & 1 & 0 & 0 & 4 \\ 0 & 2 & 0 & 1 & 0 & 12 \\ 3 & 2 & 0 & 0 & 1 & 18 \end{array}\right].$$

Obtenga esta expresión mediante la combinación de las operaciones algebraicas (en forma matricial) que afectan el renglón 0, en las iteraciones 1 y 2.

5.3-7. La mayor parte de la descripción de la idea fundamental que se presentó en la sección 5.3 supone que el problema se encuentra en nuestra forma estándar. Ahora considere otras formas que requieren los ajustes adicionales del paso inicial expuestos en la sección 4.6, inclusive el uso de variables artificiales y el método de la gran M cuando sea apropiado. Describa los ajustes que deben hacerse a la idea fundamental para:

a) Restricciones de igualdad

b) Restricciones funcionales de la forma $\geq$

c) Lados derechos negativos

d) Variables que pueden tomar valores negativos (sin cota inferior)

5.3-8. Reconsidere el problema 4.6-5. Utilice variables artificiales y el método de la gran M para construir la primera tabla símplex completa para el método símplex y después identifique las columnas que formarán $\mathbf{S}^*$ para aplicar la idea fundamental en la tabla símplex final. Explique por qué éstas son las columnas apropiadas.

5.3-9. Considere el siguiente problema.

$$\text{Minimizar} \qquad Z = 2x_1 + 3x_2 + 2x_3,$$

sujeta a

$$\begin{aligned} x_1 + 4x_2 + 2x_3 &\geq 8 \\ 3x_1 + 2x_2 + 2x_3 &\geq 6 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0.$$

Sean x_4 y x_6 las respectivas variables de superávit de las dos primeras restricciones. Sean $\bar{x}_5$ y $\bar{x}_7$ las variables artificiales correspondientes. Después de hacer los ajustes descritos en la sección 4.6 para esta forma de modelo cuando se usa el método de la gran M, la tabla símplex inicial lista para aplicar el método símplex es la siguiente:

Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	$\bar{x}_5$	x_6	$\bar{x}_7$	Lado derecho
Z	(0)	-1	$-4M+2$	$-6M+3$	$-2M+2$	M	0	M	0	$-14M$
$\bar{x}_5$	(1)	0	1	4	2	-1	1	0	0	8
$\bar{x}_7$	(2)	0	3	2	0	0	0	-1	1	6

Después de aplicar el método símplex, una parte de la tabla símplex final es la que sigue:

Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	$\bar{x}_5$	x_6	$\bar{x}_7$	Lado derecho
Z	(0)	-1					$M - 0.5$		$M - 0.5$	
x_2	(1)	0					0.3		-0.1	
x_1	(2)	0					-0.2		0.4	

a) Con base en la tabla símplex anterior, use la idea fundamental que se presentó en la sección 5.3 para identificar los números que faltan en esa tabla símplex final. Muestre sus cálculos.

b) Examine la lógica matemática que se presentó en la sección 5.3 para validar la idea fundamental (vea las ecuaciones $\mathbf{T}^* = \mathbf{MT}$ y $\mathbf{t}^* = \mathbf{t} + \mathbf{vT}$ y los desarrollos subsecuentes de $\mathbf{M}$ y $\mathbf{v}$). Esta lógica supone que el modelo original se ajusta a nuestra forma estándar mientras que este problema no lo hace. Muestre cómo, con ajustes menores, se puede aplicar la misma lógica a este problema, donde $\mathbf{t}$ es el renglón 0 y $\mathbf{T}$ está formada por los renglones 1 y 2 de la tabla símplex inicial. Obtenga $\mathbf{M}$ y $\mathbf{v}$ de este problema.

c) Al aplicar la ecuación $\mathbf{t}^* = \mathbf{t} + \mathbf{vT}$ una alternativa es usar $\mathbf{t} = [2, 3, 2, 0, M, 0, M, 0]$, que es el renglón 0 *preliminar* antes de eliminar en forma algebraica los coeficientes distintos de cero de las variables básicas iniciales $\bar{x}_5$ y $\bar{x}_7$. Repita el inciso *b*) para esta ecuación con esta nueva $\mathbf{t}$. Después de obtener la nueva $\mathbf{v}$, demuestre que esta ecuación conduce al mismo renglón 0 final que la ecuación que obtuvo en el inciso *b*).

d) Identifique las ecuaciones de definición de la solución FEV correspondiente a la solución BF óptima de la tabla símplex final.

5.3-10. Considere el siguiente problema.

$$\text{Maximizar} \qquad Z = 3x_1 + 7x_2 + 2x_3,$$

sujeta a

$$-2x_1 + 2x_2 + x_3 \leq 10$$
$$3x_1 + x_2 - x_3 \leq 20$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0.$$

Se conoce el hecho de que las variables básicas de la solución óptima son x_1 y x_3.

a) Introduzca las variables de holgura y después utilice la información que se proporcionó para encontrar la solución óptima en forma directa mediante la eliminación gaussiana.

b) Amplíe el trabajo del inciso *a*) para encontrar los precios sombra.

c) Utilice la información que se proporcionó para identificar las ecuaciones de definición de la solución FEV óptima y después resuelva estas ecuaciones para obtener la solución óptima.

d) Construya la matriz base $\mathbf{B}$ de la solución BF óptima, invierta $\mathbf{B}$ manualmente y después use esta $\mathbf{B}^{-1}$ para obtener la solución óptima y los precios sombra $\mathbf{y}^*$. Luego aplique la prueba de optimalidad del método símplex revisado para verificar que esta solución es óptima.

e) Dados $\mathbf{B}^{-1}$ y $\mathbf{y}^*$ del inciso *d*), utilice la idea fundamental presentada en la sección 5.3 para construir la tabla símplex final.

5.4-1. Considere el modelo que se presentó en el problema 5.2-2. Sean x_6 y x_7 las variables de holgura para la primera y segunda restricciones, respectivamente. Se le proporciona la información de que x_2 es a variable básica entrante y x_7 la variable básica saliente de la primera iteración del método símplex y, por lo tanto, x_4 es la variable básica de ingreso y x_6 es la variable básica de egreso de la segunda iteración (final). Utilice el procedimiento que se presentó en la sección 5.4 para actualizar $\mathbf{B}^{-1}$ de una iteración a la siguiente con el fin de encontrar $\mathbf{B}^{-1}$ después de la primera iteración y, luego, después de la segunda.

I **5.4-2.*** Utilice el método símplex revisado paso a paso para resolver el modelo que se muestra en el problema 4.3-4.

I **5.4-3.** Utilice el método símplex revisado paso a paso para resolver el modelo que se muestra en el problema 4.7-5.

I **5.4-4.** Utilice el método símplex revisado paso a paso para resolver el modelo que se muestra en el problema 3.1-6.

CAPÍTULO 6

Teoría de la dualidad y análisis de sensibilidad

Uno de los descubrimientos más importantes durante el desarrollo inicial de la programación lineal fue el concepto de dualidad y sus importantes ramificaciones. Este descubrimiento reveló que, asociado a todo problema de programación lineal, existe otro problema lineal llamado **dual**. Desde distintos puntos de vista las relaciones entre el problema dual y el original (llamado **primal**) son muy útiles. Por ejemplo, se verá que, en realidad, la solución óptima del problema dual es la que proporciona los precios sombra que se describieron en la sección 4.7. En este capítulo se presentarán muchas otras aplicaciones valiosas de la teoría de la dualidad.

Una de las aplicaciones más importantes de esta teoría es la interpretación y realización del *análisis de sensibilidad*. Como se mencionó en las secciones 2.3, 3.3 y 4.7, el análisis de sensibilidad es una parte esencial de casi todos los estudios de programación lineal. Debido a que la mayoría de los valores de los parámetros que se emplean en el modelo original son sólo *estimaciones* de las condiciones futuras, es necesario investigar el efecto que tendrían sobre la solución óptima en caso de que prevalecieran otras condiciones. Aún más, ciertos valores de estos parámetros (como la cantidad de recursos) pueden representar decisiones administrativas, en cuyo caso su elección debe ser el aspecto principal de la investigación, el cual se puede estudiar a través del análisis de sensibilidad.

Para lograr una mayor claridad, en las tres primeras secciones se presenta la teoría de dualidad bajo el supuesto de que el problema *primal* de programación lineal se presenta en *nuestra forma estándar* (pero sin la restricción de que los valores de b_i deban ser positivos). Más adelante, en la sección 6.4 se analizan otras formas. El capítulo comienza con una introducción a la esencia de la teoría de la dualidad y sus aplicaciones. Después se describe la interpretación económica del problema dual (sección 6.2) y se profundiza en las relaciones entre el problema primal y el dual (sección 6.3). En la sección 6.5 se hace hincapié en la importancia de la teoría de la dualidad para el análisis de sensibilidad, cuyo procedimiento esencial (que se basa en la idea fundamental de la sección 5.3) se resume en la sección 6.6 y se ejemplifica en la sección 6.7. En la sección 6.8 se estudia cómo usar las hojas de cálculo para realizar el análisis de sensibilidad de un modo directo. (Si usted no cuenta con mucho tiempo para dedicarle a este capítulo, puede leer para sí mismo la sección 6.8 y encontrar una breve introducción al análisis de sensibilidad.)

6.1 ESENCIA DE LA TEORÍA DE LA DUALIDAD

Dada nuestra forma estándar para el *problema primal*, que se presenta a la izquierda (quizá después de hacer conversiones cuando provienen de otras formas), su *problema dual* tiene la forma que se muestra a la derecha.

Problema primal	*Problema dual*
Maximizar $Z = \sum_{j=1}^{n} c_j x_j$, sujeta a $\sum_{j=1}^{n} a_{ij} x_j \le b_i$, para $i = 1, 2, \ldots, m$ y $x_j \ge 0$, para $j = 1, 2, \ldots, n$.	Minimizar $W = \sum_{i=1}^{m} b_i y_i$, sujeta a $\sum_{i=1}^{m} a_{ij} y_i \ge c_j$, para $j = 1, 2, \ldots, n$ y $y_i \ge 0$, para $i = 1, 2, \ldots, m$.

En consecuencia, con el problema primal en la forma de *maximización*, el problema dual está en la forma de *minimización*. Aún más, el problema dual usa exactamente los mismos *parámetros* que el problema primal, pero en diferentes lugares, tal como se resume a continuación.

1. Los coeficientes de la función objetivo del problema primal son los *lados derechos* de las restricciones funcionales del problema dual.
2. Los lados derechos de las restricciones funcionales del problema primal son los coeficientes de la función objetivo del problema dual.
3. Los coeficientes de una variable de las restricciones funcionales del problema primal son los coeficientes de una restricción funcional del problema dual.

Para recalcar esta comparación, observe ahora estos mismos problemas en la notación matricial (tal como se introdujo al principio de la sección 5.2), donde $\mathbf{c}$ y $\mathbf{y} = [y_1, y_2, \ldots, y_m]$ son vectores renglón, pero $\mathbf{b}$ y $\mathbf{x}$ son vectores columna.

Problema primal	*Problema dual*
Maximizar $Z = \mathbf{cx}$, sujeta a $\mathbf{Ax} \le \mathbf{b}$ y $\mathbf{x} \ge \mathbf{0}$.	Minimizar $\mathbf{W} = \mathbf{yb}$, sujeta a $\mathbf{yA} \ge \mathbf{c}$ y $\mathbf{y} \ge \mathbf{0}$.

A manera de ilustración, en la tabla 6.1 se muestran los problemas primal y dual del ejemplo de la Wyndor Glass Co., de la sección 3.1, tanto en forma algebraica como matricial.

La **tabla primal-dual** para programación lineal (vea la tabla 6.2) ayuda también a subrayar la correspondencia entre los dos problemas. Muestra todos los parámetros de programación lineal (las a_{ij}, b_i y c_j) y cómo se usan para construir los dos problemas. Todos los encabezados del problema primal están en posición horizontal, mientras que los del problema dual se leen al dar un cuarto de vuelta al libro. En el caso del problema primal, cada *columna* (excepto la columna del lado derecho) proporciona los coeficientes de una sola variable de las respectivas restricciones y de la función objetivo, mientras que cada *renglón* (excepto el último) da los parámetros de una restricción. En el problema dual, cada *renglón* (excepto los coeficientes del lado derecho) da los coeficientes de cada variable de las restricciones respectivas y de la función objetivo, mientras que cada *columna* (excepto la del lado derecho) proporciona los parámetros para cada restricción. Además, la columna del lado derecho da los lados derechos del problema primal y los coeficientes de la función objetivo del problema dual, mientras que el último renglón proporciona los coeficientes de la función objetivo del problema primal y los lados derechos del problema dual.

En consecuencia, ahora se tienen las siguientes relaciones generales entre los problemas primal y dual.

1. Los parámetros de una *restricción* (funcional) en cualquier problema son los coeficientes de una *variable* en el otro.

TABLA 6.1 Problemas primal y dual del ejemplo de la Wyndor Glass Co.

Problema primal en forma algebraica

$$\begin{aligned}
&\text{Maximizar} \quad Z = 3x_1 + 5x_2,\\
&\text{sujeta a}\\
&\quad x_1 \leq 4\\
&\quad 2x_2 \leq 12\\
&\quad 3x_1 + 2x_2 \leq 18\\
&\text{y} \quad x_1 \geq 0, \quad x_2 \geq 0.
\end{aligned}$$

Problema dual en forma algebraica

$$\begin{aligned}
&\text{Minimizar} \quad W = 4y_1 + 12y_2 + 18y_3,\\
&\text{sujeta a}\\
&\quad y_1 + 3y_3 \geq 3\\
&\quad 2y_2 + 2y_3 \geq 5\\
&\text{y}\\
&\quad y_1 \geq 0, \quad y_2 \geq 0, \quad y_3 \geq 0.
\end{aligned}$$

Problema primal en forma de matriz

$$\text{Maximizar} \quad Z = [3,\ 5]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

sujeta a

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 4 \\ 12 \\ 18 \end{bmatrix}$$

y

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Problema dual en forma de matriz

$$\text{Minimizar} \quad W = [y_1, y_2, y_3]\begin{bmatrix} 4 \\ 12 \\ 18 \end{bmatrix}$$

sujeta a

$$[y_1,\ y_2,\ y_3]\begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 3 & 2 \end{bmatrix} \geq [3,\ 5]$$

y

$$[y_1,\ y_2,\ y_3] \geq [0,\ 0,\ 0].$$

TABLA 6.2 Tabla primal-dual para programación lineal ilustrada por el ejemplo de la Wyndor Glass Co.

a) Caso general

			Problema primal					
			Coeficiente de:				Lado derecho	
			x_1	x_2	$\cdots$	x_n		
Problema dual	Coeficiente de:	y_1	a_{11}	a_{12}	$\cdots$	a_{1n}	$\leq b_1$	Coeficientes de la función objetivo (minimizar)
		y_2	a_{21}	a_{22}	$\cdots$	a_{2n}	$\leq b_2$	
		$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$	
		y_m	a_{m1}	a_{m2}	$\cdots$	a_{mn}	$\leq b_m$	
	Lado derecho		VI	VI	$\cdots$	VI		
			c_1	c_2	$\cdots$	c_n		
			Coeficientes de la función objetivo (maximizar)					

b) Ejemplo de la Wyndor Glass Co.

	x_1	x_2	
y_1	1	0	≤ 4
y_2	0	2	≤ 12
y_3	3	2	≤ 18
	VI	VI	
	3	5	

2. Los coeficientes de la *función objetivo* en un problema son los valores del *lado derecho* en el otro.

En consecuencia, existe una correspondencia directa entre los elementos de los dos problemas, como se resume en la tabla 6.3. Esta correspondencia es la clave de algunas aplicaciones de la teoría de la dualidad entre las que se encuentra el análisis de sensibilidad.

En la sección Worked Examples del sitio en internet de este libro se proporciona **otro ejemplo** del uso de la tabla primal-dual para construir el problema dual de un modelo de programación lineal.

Origen del problema dual

La teoría de la dualidad se basa de manera directa en la *idea fundamental* (en particular con respecto al renglón 0) que se presentó en la sección 5.3. Para ver por qué, se continuará con el uso de la notación que se introdujo en la tabla 5.9 para el renglón 0 de la tabla símplex final, pero se reemplazará Z^* por W^* y se quitarán los asteriscos de $\mathbf{z}^*$ y $\mathbf{y}^*$ cuando se haga referencia a *cualquier* tabla símplex. Entonces, en una iteración dada del método símplex en el problema primal los números actuales del renglón 0 se denotan como se muestra en la tabla símplex (parcial) de la tabla 6.4. En el caso de los coeficientes de $x_1, x_2, \ldots, x_n$, recuerde que $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ denota el vector que agrega el método símplex al vector de coeficientes *iniciales*, $-\mathbf{c}$, en el proceso de obtener la tabla símplex actual. (No confunda $\mathbf{z}$ con el valor de la función objetivo Z.) De manera similar, como los coeficientes *iniciales* de $x_{n+1}, x_{n+2}, \ldots, x_{n+m}$ del renglón 0 son todos iguales a 0, $\mathbf{y} = (y_1, y_2, \ldots, y_m)$ denota el vector que agrega el método símplex a estos coeficientes. Recuerde también [vea la ecuación (1) en la subsección del "Resumen matemático" de la sección 5.3] que la idea fundamental conducía a las siguientes relaciones entre las cantidades y los parámetros del modelo original:

$$W = \mathbf{yb} = \sum_{i=1}^{m} b_i y_i,$$

$$\mathbf{z} = \mathbf{yA}, \qquad \text{de manera que} \qquad z_j = \sum_{i=1}^{m} a_{ij} y_i, \qquad \text{para } j = 1, 2, \ldots, n.$$

Para ilustrar estas relaciones con el ejemplo de la Wyndor, la primera ecuación resulta $W = 4y_1 + 12y_2 + 18y_3$, que es justo la función objetivo del problema dual que se presenta en la esquina superior derecha de la tabla 6.1. El segundo conjunto de ecuaciones resulta $z_1 = y_1 + 3y_3$ y $z_2 = 2y_2 + 2y_3$, que son los lados izquierdos de las restricciones funcionales de este problema dual. Así, si se restan los lados derechos de estas restricciones tipo $\geq$ entonces ($c_1 = 3$ y $c_2 = 5$), $(z_1 - c_1)$ y $(z_2 - c_2)$ se pueden interpretar como las *variables de superávit* de estas restricciones funcionales.

TABLA 6.3 Correspondencia entre los elementos de los problemas primal y dual

Un problema		Otro problema
Restricción i	⟷	Variable i
Función objetivo	⟷	Lados derechos

TABLA 6.4 Notación de los elementos del renglón 0 en la tabla símplex

Iteración	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	$\cdots$	x_n	x_{n+1}	x_{n+2}	$\cdots$	x_{n+m}	Lado derecho
Cualquiera	Z	(0)	1	$z_1 - c_1$	$z_2 - c_2$	$\cdots$	$z_n - c_n$	y_1	y_2	$\cdots$	y_m	W

La clave ahora es expresar lo que el método símplex trata de lograr (de acuerdo con la prueba de optimalidad) en términos de estos símbolos. En particular, busca un conjunto de variables básicas y la solución BF correspondiente, tal que *todos* los coeficientes del renglón 0 sean *no negativos*. Después se detiene con esta solución óptima. Con la notación de la tabla 6.4, esto se expresa simbólicamente de la siguiente forma:

Condición de optimalidad:

$z_j - c_j \geq 0 \quad$ para $j = 1, 2, \ldots, n,$

$y_i \geq 0 \quad$ para $i = 1, 2, \ldots, m.$

Después de sustituir la expresión anterior de z_j, la condición de optimalidad dice que el método símplex se puede interpretar como la búsqueda de los valores de $y_1, y_2, \ldots, y_m$ tales que

$$W = \sum_{i=1}^{m} b_i y_i,$$

sujeta a

$$\sum_{i=1}^{m} a_{ij} y_i \geq c_j, \quad \text{para } j = 1, 2, \ldots, n$$

y

$$y_i \geq 0, \quad \text{para } i = 1, 2, \ldots, m.$$

Pero, excepto por la falta de un objetivo para la función W, este problema es precisamente el *problema dual*. Para completar la formulación, se explorará cuál debe ser ese objetivo perdido.

Como W es sencillamente el valor de Z, y como el objetivo del problema primal es maximizar Z, una reacción natural sería que W también se maximizara. Sin embargo, esto no es correcto debido a una razón bastante sutil: las únicas soluciones *factibles* para este nuevo problema son aquellas que satisfacen la condición de *optimalidad* del problema primal. Por tanto, *sólo* la solución óptima para el problema primal corresponde a la solución factible de este nuevo problema. En consecuencia, el valor óptimo de Z en el problema primal es el valor *mínimo* factible de W en el nuevo problema, de manera que W debe minimizarse. (Las relaciones que se desarrollan en la sección 6.3 proporcionan la justificación completa de esta conclusión.) Si se agrega este objetivo de minimizar W, se obtiene el problema dual *completo*.

En consecuencia, el problema dual se puede ver como otra forma de establecer, en términos de programación lineal, la meta del método símplex, que es alcanzar una solución para el problema primal que *satisfaga la prueba de optimalidad*. *Antes* de alcanzar esta meta, la $\mathbf{y}$ correspondiente del renglón 0 (los coeficientes de las variables de holgura) de la tabla símplex final debe ser *no factible* para el problema dual. Sin embargo, después de lograr la meta, la $\mathbf{y}$ correspondiente debe ser una *solución óptima* (indicada por $\mathbf{y}^*$) del *problema dual*, puesto que se trata de una solución factible que adquiere el valor factible más pequeño de W. Esta solución óptima $(y_1^*, y_2^*, \ldots, y_m^*)$ proporciona los precios sombra del problema primal que se describieron en la sección 4.7. Aún más, esta W óptima no es otra cosa que el valor óptimo de Z, de manera que los *valores de la función objetivo son iguales* en los dos problemas. Este hecho también significa que $\mathbf{cx} \leq \mathbf{yb}$ para cualesquiera $\mathbf{x}$ y $\mathbf{y}$ *factibles* de los problemas primal y dual, respectivamente.

Como ilustración, en el lado izquierdo de la tabla 6.5 se muestra el renglón 0 de las iteraciones respectivas cuando se aplica el método símplex al ejemplo de la Wyndor Glass Co. En cada caso, se dividió el renglón 0 en tres partes: los coeficientes de las variables originales (x_1, x_2), los coeficientes de las variables de holgura (x_3, x_4, x_5) y el lado derecho (valor de Z). Dado que los coeficientes de las variables de holgura proporcionan los valores que corresponden a las variables duales (y_1, y_2, y_3), cada renglón 0 identifica una solución correspondiente para el problema dual, como se muestra en las columnas de y_1, y_2 y y_3 de la tabla 6.5. Para interpretar las dos columnas siguientes, recuerde que $(z_1 - c_1)$ y $(z_2 - c_2)$ son las variables de superávit de las restricciones funcionales del problema dual, de modo que el problema dual completo después de aumentar estas variables de superávit es

Minimizar $\quad W = 4y_1 + 12y_2 + 18y_3,$

sujeta a

$$\begin{aligned} y_1 \quad + 3y_3 - (z_1 - c_1) &= 3 \\ 2y_2 + 2y_3 - (z_2 - c_2) &= 5 \end{aligned}$$

y

$$y_1 \geq 0, \qquad y_2 \geq 0, \qquad y_3 \geq 0.$$

Entonces, si se usan los números de las columnas de y_1, y_2 y y_3 los valores de estas variables de superávit se pueden calcular como

$$\begin{aligned} z_1 - c_1 &= y_1 + 3y_3 - 3, \\ z_2 - c_2 &= 2y_2 + 2y_3 - 5. \end{aligned}$$

De esta forma, un valor negativo de cualquier variable de superávit indica que se viola la restricción correspondiente. En la columna de la derecha también se incluye el valor calculado de la función objetivo dual $W = 4y_1 + 12y_2 + 18y_3$.

Como se mostró en la tabla 6.4, *todas* estas cantidades de la derecha del renglón 0 se identifican en este renglón sin necesidad de hacer nuevos cálculos. En particular, observe en la tabla 6.5 cómo *cada* número que se obtuvo para el problema dual ya aparece en el renglón 0 en la posición indicada de la tabla 6.4.

En el caso del renglón 0 inicial, en la tabla 6.5 se muestra que la solución dual correspondiente, $(y_1, y_2, y_3) = (0, 0, 0)$, es no factible, puesto que las dos variables de superávit son negativas. La primera iteración logra eliminar uno de estos dos valores negativos, pero no el otro. Después de dos iteraciones se satisface la prueba de optimalidad del problema primal debido a que todas las variables duales y las de superávit son no negativas. Esta solución dual $(y_1^*, y_2^*, y_3^*) = (0, \frac{3}{2}, 1)$ es óptima (como puede verificarse si se aplica el método símplex de manera directa al problema dual), por lo cual el valor óptimo de Z y de W es $Z^* = 36 = W^*$.

Resumen de las relaciones primal-dual

A continuación se presenta un resumen de las importantes relaciones entre los problemas primal y dual que se acaban de descubrir.

Propiedad de dualidad débil: Si **x** es una solución factible para el problema primal y **y** es una solución factible para el problema dual, entonces

$$\mathbf{cx} \leq \mathbf{yb}.$$

Por ejemplo, en el problema de la Wyndor Glass Co., una solución factible es $x_1 = 3$, $x_2 = 3$, lo que conduce a $Z = \mathbf{cx} = 24$, y una solución factible del problema dual es $y_1 = 1$, $y_2 = 1$, $y_3 = 2$, que resulta en un valor más grande de la función objetivo $W = \mathbf{yb} = 52$. Éstos son sólo ejemplos de las soluciones factibles para los dos problemas. Para *cualquier* par de soluciones factibles, esta desigualdad debe cumplirse debido a que el valor factible *máximo* de $Z = \mathbf{cx}$ (36) es *igual* al valor factible *mínimo* de la función objetivo dual $W = \mathbf{yb}$, que es la siguiente propiedad.

TABLA 6.5 El renglón 0 y su correspondiente solución dual para cada iteración del ejemplo de la Wyndor Glass Co.

	Problema primal	Problema dual					
Iteración	**Renglón 0**	y_1	y_2	y_3	$z_1 - c_1$	$z_2 - c_2$	**W**
0	$[-3, \ -5 \ \vdots \ 0, \ 0, \ 0 \ \vdots \ 0]$	0	0	0	−3	−5	0
1	$[-3, \ 0 \ \vdots \ 0, \ \frac{5}{2}, \ 0 \ \vdots \ 30]$	0	$\frac{5}{2}$	0	−3	0	30
2	$[\ 0, \ 0 \ \vdots \ 0, \ \frac{3}{2}, \ 1 \ \vdots \ 36]$	0	$\frac{3}{2}$	1	0	0	36

Propiedad de dualidad fuerte: Si $\mathbf{x}^*$ es una solución óptima para el problema primal y $\mathbf{y}^*$ es una solución óptima para el problema dual, entonces

$$\mathbf{cx}^* = \mathbf{y}^*\mathbf{b}.$$

Estas dos propiedades implican que $\mathbf{cx} < \mathbf{yb}$ para soluciones factibles si una o ambas son *no óptimas* para sus problemas respectivos, mientras que la igualdad se cumple cuando ambas son óptimas.

La *propiedad de dualidad débil* describe la relación entre cualquier par de soluciones de los problemas primal y dual en donde *ambas* soluciones son *factibles* para sus problemas respectivos. En cada iteración, el método símplex encuentra un par específico de soluciones para los dos problemas, donde la solución del problema primal es factible pero la del dual es *no factible* (excepto en la última iteración). La siguiente propiedad describe esta situación y la relación entre este par de soluciones.

Propiedad de soluciones complementarias: En cada iteración, el método símplex identifica de manera simultánea una solución FEV, $\mathbf{x}$, para el problema primal y una **solución complementaria**, $\mathbf{y}$, para el problema dual (que se encuentra en el renglón 0, como los coeficientes de las variables de holgura), donde

$$\mathbf{cx} = \mathbf{yb}.$$

Si $\mathbf{x}$ *no es óptima* para el problema primal, entonces $\mathbf{y}$ *no es factible* para el problema dual.

Para ilustrar esta propiedad, después de una iteración en el problema de la Wyndor Glass Co., $x_1 = 0$, $x_2 = 6$, y $y_1 = 0$, $y_2 = \frac{5}{2}$, $y_3 = 0$, con $\mathbf{cx} = 30 = \mathbf{yb}$. Esta $\mathbf{x}$ es factible para el problema primal, pero $\mathbf{y}$ es no factible para el dual (dado que viola la restricción, $y_1 + 3y_3 \geq 3$).

La propiedad de soluciones complementarias también se cumple en la iteración final del método símplex, donde se encuentra una solución óptima para el problema primal. Sin embargo, se puede decir más sobre la solución complementaria y en este caso, como se presenta en la siguiente propiedad.

Propiedad de soluciones complementarias óptimas: Al final de cada iteración, el método símplex identifica de manera simultánea una solución óptima $\mathbf{x}^*$ para el problema primal y una **solución óptima complementaria** $\mathbf{y}^*$ para el problema dual (que se encuentra en el renglón 0 como los coeficientes de las variables de holgura), donde

$$\mathbf{cx}^* = \mathbf{y}^*\mathbf{b}.$$

Los valores de y_i^* son los precios sombra para el problema primal.

En el ejemplo, de la iteración final se obtiene $x_1^* = 2$, $x_2^* = 6$, y $y_1^* = 0$, $y_2^* = \frac{3}{2}$, $y_3^* = 1$, con $\mathbf{cx}^* = 36 = \mathbf{y}^*\mathbf{b}$.

En la sección 6.3 se analizarán con más detenimiento algunas de estas propiedades. Se verá que la propiedad de soluciones complementarias se puede extender mucho más. En particular, después de introducir las variables de holgura y de superávit en ambos problemas, toda solución *básica* del problema primal tiene una solución *básica* complementaria del problema dual. Ya se vio que el método símplex identifica los valores de las variables de superávit del problema dual como $z_j - c_j$ en la tabla 6.4. Este resultado conduce después a una *propiedad adicional de holgura complementaria* que relaciona las variables básicas de un problema con las no básicas del otro (tablas 6.7 y 6.8), tema que después se estudiará más profundamente.

En la sección 6.4, después de describir cómo se construye el problema dual cuando el problema primal no se encuentra en nuestra forma estándar, se analizará otra propiedad útil que se resume de la forma siguiente:

Propiedad de simetría: En el caso de *cualquier* problema primal y su problema dual, las relaciones entre ellos deben ser *simétricas* debido a que el dual de este problema dual es este problema primal.

En consecuencia, todas las propiedades anteriores se cumplen sin que importe a cuál de los dos problemas se le llame problema primal. (La dirección de la desigualdad de la propiedad de dua-

lidad débil requiere que el problema primal se exprese o reexprese en la forma de maximización y el problema dual en la forma de minimización.) Por tanto, el método símplex se puede aplicar a cualquiera de los dos problemas e identificará al mismo tiempo las soluciones complementarias (y en última instancia una solución complementaria óptima) para el otro problema.

Hasta este punto se han estudiado las relaciones entre las soluciones *factibles* u *óptimas* para el problema primal y las soluciones correspondientes para el problema dual. Sin embargo, es posible que el problema primal (o el dual) no tenga *soluciones factibles* o bien tenga soluciones factibles pero *no una solución óptima* (debido a que la función objetivo no esté acotada). La última propiedad resume las relaciones primal-dual de todas estas posibilidades.

Teorema de la dualidad: Las siguientes son las únicas relaciones posibles entre los problemas primal y dual.

1. Si un problema tiene *soluciones factibles* y una función objetivo acotada (y, por ende, una solución óptima), entonces ocurre lo mismo con el otro problema, de manera que se aplican tanto la propiedad de dualidad débil como la fuerte.
2. Si uno de los problemas tiene *soluciones factibles* y una función objetivo no acotada (es decir, *no tiene solución óptima*), entonces el otro problema *no tiene soluciones factibles*.
3. Si un problema *no tiene soluciones factibles*, entonces el otro problema *no tiene soluciones factibles* o bien la función objetivo es *no acotada*.

Aplicaciones

Como acaba de establecerse, una aplicación importante de la teoría de la dualidad es que puede resolverse el problema *dual* directamente con el método símplex, a fin de identificar una solución óptima para el problema primal. En la sección 4.8 se explicó que el número de restricciones funcionales afecta mucho más el esfuerzo computacional del método símplex que el número de variables. Si $m > n$, el problema dual tiene menos restricciones funcionales (n) que el problema primal (m), entonces es probable lograr una reducción considerable del esfuerzo computacional si se aplica el método símplex en forma directa al problema dual en lugar de aplicarlo al primal.

Las *propiedades de dualidad débil* y *fuerte* describen las relaciones clave entre los problemas primal y dual. Una aplicación útil es la evaluación de una solución propuesta para el problema primal. Por ejemplo, suponga que $\mathbf{x}$ es una solución factible que se ha propuesto para su implantación, y que se ha encontrado, por inspección del dual, una solución factible $\mathbf{y}$ tal que $\mathbf{cx} = \mathbf{yb}$. En este caso, $\mathbf{x}$ debe ser *óptima* sin que sea necesario aplicar el método símplex. Aun si $\mathbf{cx} < \mathbf{yb}$, de todas maneras $\mathbf{yb}$ proporciona una cota superior sobre el valor óptimo de Z, de manera que si $\mathbf{yb} - \mathbf{cx}$ es pequeño, los factores intangibles que favorecen a $\mathbf{x}$ pueden conducir a que se elija esta solución sin dedicarle más esfuerzo.

En la sección 7.1 se presenta el método símplex dual, que es una de las aplicaciones más importantes de la propiedad de soluciones complementarias. Este algoritmo opera sobre el problema primal como si se aplicara el método símplex en forma simultánea al problema dual, lo que puede realizarse gracias a esta propiedad. Debido a que en la tabla símplex se invierten los papeles del renglón 0 y la columna del lado derecho, el método símplex dual requiere que el renglón 0 *comience y permanezca no negativo* mientras que el lado derecho *comienza* con algunos valores *negativos* (en las iteraciones siguientes se procura alcanzar la no negatividad en el lado derecho). En consecuencia, algunas veces se usa este algoritmo porque es más conveniente establecer la tabla inicial en esta forma que en la que requiere el método símplex. Aún más, con frecuencia se usa para la reoptimización (tema que se presentó en la sección 4.7), porque algunos cambios en el modelo original conducen a una tabla final revisada que se ajusta a esta forma. Esta situación es común en cierto tipo de análisis de sensibilidad, como se verá más adelante en este capítulo.

En términos generales, la teoría de dualidad tiene un papel central en el análisis de sensibilidad. Su importancia es el tema de la sección 6.5.

Otra aplicación importante es su uso en la interpretación económica del problema dual y la visión que se obtiene para el análisis del problema primal. Ya se vio un ejemplo cuando se estudiaron los precios sombra en la sección 4.7. En la sección 6.2 se describe la forma en que esta interpretación se extiende a todo el problema dual y después al método símplex.

6.2 INTERPRETACIÓN ECONÓMICA DE LA DUALIDAD

La interpretación económica de la dualidad se basa de manera directa en la interpretación más frecuente del problema primal (problema de programación lineal en nuestra forma estándar), tema que se estudió en la sección 3.2. Para refrescar la memoria, se ha resumido esta interpretación del problema primal en la tabla 6.6.

Interpretación del problema dual

Para ver la forma en que esta interpretación del problema primal conduce a una interpretación económica del problema dual,[1] observe en la tabla 6.4 que W es el valor de Z (utilidad total) en la iteración actual. Como

$$W = b_1y_1 + b_2y_2 + \cdots + b_my_m,$$

cada b_iy_i puede interpretarse como la *contribución a la ganancia* por disponer de b_i unidades del recurso i en el problema primal. Así,

> la variable y_i se interpreta como la contribución a la ganancia por unidad del recurso i ($i = 1, 2, \ldots, m$), cuando se usa el conjunto actual de variables básicas para obtener la solución primal.

En otras palabras, los valores de y_i (o los valores de y_i^* en la solución óptima) no son otra cosa que los **precios sombra** que se presentaron en la sección 4.7.

Por ejemplo, cuando la iteración 2 del método símplex encuentra la solución óptima para el problema de la Wyndor, también encuentra que los valores óptimos de las variables duales (como se muestra en el último renglón de la tabla 6.5) son $y_1^* = 0$, $y_2^* = \frac{3}{2}$ y $y_3^* = 1$. Éstos son precisamente los precios sombra que se encontraron en la sección 4.7 para este problema mediante el método gráfico. Recuerde que los recursos son las capacidades de producción de las tres plantas disponibles para los dos nuevos productos bajo consideración, de modo que b_i es el número de horas semanales de producción disponibles en la planta i para estos nuevos productos, donde $i = 1, 2, 3$. Como se vio en la sección 4.7, los precios sombra indican que un incremento individual de 1 en cualquier b_i aumentará en y_i^* el valor óptimo de la función objetivo (ganancia total semanal en miles de dólares). Así y_i^* se puede interpretar como la contribución a la utilidad por unidad del recurso i al usar la solución óptima.

Esta interpretación de las variables duales conduce a la interpretación del problema dual completo. En especial, como cada unidad de la actividad y del problema primal consume a_{ij} unidades del recurso i,

> $\sum_{i=1}^{m} a_{ij}\, y_i$ se interpreta como la contribución actual a la utilidad de esa mezcla de recursos que se consumiría si se usara 1 unidad de la actividad ($j = 1, 2, \ldots, n$).

En el caso del problema de la Wyndor, 1 unidad de la actividad j corresponde a producir 1 lote del producto j por semana, donde $j = 1, 2$. La mezcla de recursos consumida al producir 1 lote del producto 1 es 1 hora de producción de la planta 1 y 3 horas de la planta 3. La mezcla correspondiente por lote del producto 2 es 2 horas de cada una de las plantas 2 y 3. En consecuencia, y_1 +

TABLA 6.6 Interpretación económica del problema primal

Cantidad	Interpretación
x_j	Nivel de la actividad j ($j = 1, 2, \ldots, n$)
c_j	Ganancia unitaria debida a la actividad j
Z	Ganancia total debida a todas las actividades
b_i	Cantidad disponible del recurso i ($i = 1, 2, \ldots, m$)
a_{ij}	Cantidad del recurso i consumida por cada unidad de la actividad j

[1] En realidad se han propuesto varias interpretaciones, con ligeras diferencias. La que se presenta fue considerada por los autores como la más útil, puesto que también interpreta directamente lo que hace el método símplex en el problema primal.

$3y_3$ y $2y_2 + 2y_3$ se interpretan como las contribuciones actuales a la utilidad (en miles de dólares semanales) de estas mezclas respectivas de recursos por lote producido por semana de los productos respectivos.

En el caso de cada actividad j, esta misma mezcla de recursos (y más) quizá se pueda usar de otra manera, pero no debe considerarse ningún otro uso si es menos redituable que 1 unidad de la actividad j. Debido a que c_j se interpreta como la utilidad unitaria que se obtiene por la actividad j, cada restricción funcional del problema dual se interpreta de la siguiente manera:

> $\sum_{i=1}^{m} a_{ij} y_i \geq c_j$ dice que la contribución actual a la utilidad de la mezcla anterior de recursos debe ser, por lo menos, tan grande como si 1 unidad de la actividad j la utilizara; de otra manera no se llevaría a cabo la mejor utilización de estos recursos.

En el problema de la Wyndor, las ganancias unitarias (en miles de dólares por semana) son $c_1 = 3$ y $c_2 = 5$, de manera que las restricciones funcionales duales que se obtienen con esta interpretación son $y_1 + 3y_3 \geq 3$ y $2y_2 + 2y_3 \geq 5$. De manera similar, la interpretación de las restricciones de no negatividad es la siguiente:

> $y_i \geq 0$ dice que la contribución a la utilidad por parte del recurso i ($i = 1, 2, \ldots, m$) debe ser no negativa, pues de lo contrario sería mejor no utilizar este recurso.

El objetivo

$$\text{Minimizar} \qquad W = \sum_{i=1}^{m} b_i y_i$$

puede verse como la minimización del valor total implícito de los recursos consumidos por las actividades. En el problema de la Wyndor, el valor implícito total (en miles de dólares por semana) de los recursos consumidos por los dos productos es $W = 4y_1 + 12y_2 + 18y_3$.

Esta interpretación se puede perfeccionar un poco si se toma en cuenta la diferencia entre las variables básicas y las no básicas en el problema primal para cualquier solución BF dada $(x_1, x_2, \ldots, x_{n+m})$. Recuerde que las variables básicas (las únicas que pueden tomar valores distintos de cero) siempre tienen un coeficiente igual a *cero* en el renglón 0. Por tanto, si se hace referencia de nuevo a la tabla 6.4 y a la ecuación correspondiente para z_j, se observa que

$$\sum_{i=1}^{m} a_{ij} y_i = c_j, \qquad \text{si } x_j > 0 \qquad (j = 1, 2, \ldots, n),$$

$$y_i = 0, \qquad \text{si } x_{n+i} > 0 \qquad (i = 1, 2, \ldots, m).$$

(Ésta es una versión de la propiedad de holgura complementaria que se estudiará en la sección 6.3.) La interpretación económica de la primera afirmación es que siempre que una actividad j opere a un nivel estrictamente positivo ($x_j > 0$), el valor marginal de los recursos que consume *debe ser igual* (como opuesto a exceder) a la ganancia unitaria de esta actividad. La segunda afirmación indica que el valor marginal del recurso i es *cero* ($y_i = 0$) siempre y cuando las actividades ($x_{n+i} > 0$). En terminología económica, un recurso de este tipo es un "bien gratuito"; el precio de los bienes que tienen una sobredisponibilidad debe ser igual a cero por la ley de la oferta y la demanda. Este hecho justifica la interpretación de la función objetivo para el problema dual como la minimización del valor de los recursos *consumidos*, en lugar de los *asignados*.

Para ilustrar estos dos enunciados, considere la solución BF óptima (2, 6, 2, 0, 0) para el problema de la Wyndor. Las variables básicas son x_1, x_2 y x_3, de manera que sus coeficientes en el renglón 0 son iguales a cero, como se muestra en el último renglón de la tabla 6.5. Este renglón también proporciona la solución dual correspondiente: $y_1^* = 0$, $y_2^* = \frac{3}{2}$, $y_3^* = 1$, con variables de superávit $(z_1^* - c_1) = 0$ y $(z_2^* - c_2) = 0$. Como $x_1 > 0$ y $x_2 > 0$, las dos variables de superávit y los cálculos directos indican que $y_1^* + 3y_3^* = c_1 = 3$ y $2y_2^* + 2y_3^* = c_2 = 5$. Por tanto, el valor de los recursos consumidos por lote de los respectivos productos sin duda es igual a las utilidades unitarias respectivas. La variable de holgura para la restricción sobre la capacidad usada en la planta 1 es $x_3 > 0$, por lo que el valor marginal de agregar cualquier capacidad en la planta 2 sería igual a cero ($y_1^* = 0$).

Interpretación del método símplex

La interpretación del problema dual proporciona también una interpretación económica de lo que hace el método símplex en el problema dual. La *meta* del símplex es encontrar la manera de usar los recursos disponibles en la forma más redituable. Para alcanzarla, debe llegar a una solución BF que satisfaga todos los *requisitos* sobre el uso provechoso de los recursos (las restricciones del problema dual). Estos requisitos comprenden la *condición de optimalidad* del algoritmo. Para cualquier solución BF dada, los requisitos (restricciones duales) asociados con las variables básicas se satisfacen de manera automática (con la igualdad). Sin embargo, los asociados con las variables no básicas pueden o no ser satisfechos.

En particular, si una variable original x_j es no básica y por ende la actividad j no se usa, la contribución actual a la utilidad debida a esos recursos, que se requerirían para emprender cada unidad de la actividad j

$$\sum_{i=1}^{m} a_{ij}y_i$$

puede ser más pequeña, más grande o igual que la ganancia unitaria c_j que puede obtenerse de dicha actividad. Si es menor, de manera que $z_j - c_j < 0$ en el renglón 0 de la tabla símplex, entonces estos recursos se pueden usar en forma más ventajosa si se inicia con esta actividad. Si es mayor ($z_j - c_j > 0$), estos recursos ya se habrán asignado en otra parte de una forma más provechosa, por lo que no deben ser designadas a la actividad j. Si $z_j - c_j = 0$, no habrá cambio en el rendimiento al iniciar la actividad j.

De manera similar, si la variable de holgura x_{n+i} es no básica, es decir, si se usa la asignación total b_i del recurso i, entonces y_i es la contribución actual a la ganancia de este recurso sobre una base marginal. De esta forma, si $y_i < 0$, la ganancia se puede incrementar al disminuir el uso de este recurso (es decir, al aumentar x_{n+i}). Si $y_i > 0$, vale la pena continuar con el uso total de este recurso, puesto que esta decisión no afecta el rendimiento si $y_i = 0$.

Por tanto, lo que hace el método símplex es examinar todas las variables no básicas de la solución BF actual para ver cuáles pueden proporcionar un *uso más ventajoso de los recursos* cuando se las incrementa. Si *ninguna* puede, es decir, si ningún cambio o reducción factible de la asignación actual propuesta de los recursos puede aumentar la utilidad, entonces la solución actual será óptima. Si una o más variables pueden aumentarla, el método símplex selecciona aquella que, si se aumenta en una unidad, proporciona el mayor *incremento a la utilidad*. Después, el valor de esta variable (la variable básica entrante) en realidad aumenta tanto como puede hasta que los valores marginales de los recursos cambian. El resultado de este incremento es una nueva solución BF con un nuevo renglón 0 (solución dual), y se repite el proceso completo.

La interpretación económica del problema dual expande en forma considerable la capacidad para analizar el problema primal. Sin embargo, ya se vio en la sección 6.1 que esta interpretación es sólo una ramificación de las relaciones entre los dos problemas. En la sección 6.3 se profundizará en estas relaciones.

6.3 RELACIONES PRIMAL-DUAL

Como el problema dual es un problema de programación lineal, también tiene soluciones en los vértices. Aún más, al emplear la forma de igualdades del problema, estas soluciones se pueden expresar como soluciones básicas. Debido a que las restricciones funcionales tienen la forma $\geq$, la forma aumentada se obtiene al *restar* el superávit (en lugar de agregar la holgura) del lado izquierdo de cada restricción j ($j = 1, 2, \ldots, n$).[2] Este superávit es

$$z_j - c_j = \sum_{i=1}^{m} a_{ij}y_i - c_j, \qquad \text{para } j = 1, 2, \ldots, n.$$

[2] El lector puede preguntarse por qué no se introducen *variables artificiales* en estas restricciones como se estudió en la sección 4.6. La razón es que estas variables no cumplen con otro propósito que cambiar temporalmente la región factible, lo cual es conveniente para iniciar el método símplex. En este caso no se tiene interés en aplicar el método símplex al problema dual y no se quiere cambiar la región factible.

En consecuencia, $z_j - c_j$ asume el papel de *variable de superávit* de la restricción j (o su variable de holgura si la restricción se multiplica por -1). Por tanto, cuando aumenta cada solución en un vértice $(y_1, y_2, \ldots, y_m)$ se obtiene una solución básica $(y_1, y_2, \ldots, y_m, z_1 - c_1, z_2 - c_2, \ldots, z_n - c_n)$ al usar esta expresión para $z_j - c_j$. Como la forma aumentada del problema dual tiene n restricciones funcionales y n m variables, cada solución básica tiene n variables básicas y m variables no básicas. (Observe que los papeles de m y n se han invertido, como se indica en la tabla 6.3, porque las restricciones duales corresponden a las variables primales y las variables duales a las restricciones primales.)

Soluciones básicas complementarias

Una de las relaciones más importantes entre los problemas primal y dual es la correspondencia directa entre sus soluciones básicas. La clave de esta correspondencia es el renglón 0 de la tabla símplex de la solución básica primal, como se muestra en las tablas 6.4 o 6.5. Este renglón 0 se puede obtener a partir de *cualquier* solución básica primal, factible o no, para lo cual se emplean las fórmulas que se presentan en la parte inferior de la tabla 5.8.

Observe de nuevo en las tablas 6.4 y 6.5 cómo se puede leer de manera directa en el renglón 0 la solución completa del problema dual (inclusive las variables de superávit). En consecuencia, debido a su coeficiente en el renglón 0, cada variable del problema primal tiene una variable asociada en el dual; esta correspondencia se resume en la tabla 6.7, primero para cualquier problema y luego para el de la Wyndor.

Una idea clave aquí es que la solución del dual que se lee en el renglón 0 también debe ser una solución básica. Esto se debe a que cada una de las m variables básicas del problema primal necesitan tener coeficiente igual a cero en el renglón 0, lo que significa que el valor de las m variables duales asociadas debe ser cero, es decir, deben ser variables no básicas del problema dual. Los valores de las n variables (básicas) restantes serán entonces la solución simultánea del sistema de ecuaciones dado al principio de esta sección. En forma matricial, este sistema es $\mathbf{z} - \mathbf{c} = \mathbf{yA} - \mathbf{c}$, y la idea fundamental de la sección 5.3, en realidad, identifica su solución para los valores de $\mathbf{z} - \mathbf{c}$ y $\mathbf{y}$ como elementos correspondientes en el renglón 0.

Debido a la propiedad de simetría que se mencionó en la sección 6.1 (y a la asociación directa entre las variables que se muestran en la tabla 6.7), la correspondencia entre las soluciones básicas en el primal y el dual es simétrica. Aún más, un par de soluciones básicas complementarias tiene el mismo valor para la función objetivo: éste es el de W en la tabla 6.4.

Ahora se resumirán las conclusiones sobre la correspondencia entre las soluciones básicas del primal y el dual, donde la primera propiedad amplía la propiedad de soluciones complementarias de la sección 6.1 a las formas aumentadas de los dos problemas y después a cualquier solución básica (factible o no) del problema primal.

> **Propiedad de las soluciones básicas complementarias:** Cada solución *básica* del *problema primal* tiene una **solución básica complementaria** para el *problema dual*, donde los valores respectivos de la función objetivo (Z y W) son iguales. Dado el renglón 0 de la tabla símplex de la solución básica primal, la solución básica dual complementaria $(\mathbf{y}, \mathbf{z} - \mathbf{c})$ se encuentra de la forma que se muestra en la tabla 6.4.

La siguiente propiedad muestra cómo identificar las variables básicas y no básicas de esta solución básica complementaria.

TABLA 6.7 Asociación entre las variables de los problemas primal y dual

	Variable primal	Variable dual asociada
Cualquier problema	(Variable de decisión) x_j (Variable de holgura) x_{n+i}	$z_j - c_j$ (variable de superávit) $j = 1, 2, \ldots, n$ y_i (variable de decisión) $i = 1, 2, \ldots, m$
Problema de Wyndor	Variables de decisión: x_1 x_2 Variables de holgura: x_3 x_4 x_5	$z_1 - c_1$ (variables de superávit) $z_2 - c_2$ y_1 (variables de decisión) y_2 y_3

TABLA 6.8 Relaciones de holgura complementaria para soluciones básicas complementarias

Variable primal	Variable dual asociada	
Básica	No básica	(*m* variables)
No básica	Básica	(*n* variables)

TABLA 6.9 Soluciones básicas complementarias para el ejemplo de la Wyndor Glass Co.

	Problema primal			Problema dual	
Núm.	Solución básica	¿Factible?	$Z = W$	¿Factible?	Solución básica
1	(0, 0, 4, 12, 18)	Sí	0	No	(0, 0, 0, −3, −5)
2	(4, 0, 0, 12, 6)	Sí	12	No	(3, 0, 0, 0, −5)
3	(6, 0, −2, 12, 0)	No	18	No	(0, 0, 1, 0, −3)
4	(4, 3, 0, 6, 0)	Sí	27	No	$\left(-\frac{9}{2}, 0, \frac{5}{2}, 0, 0\right)$
5	(0, 6, 4, 0, 6)	Sí	30	No	$\left(0, \frac{5}{2}, 0, -3, 0\right)$
6	(2, 6, 2, 0, 0)	Sí	36	Sí	$\left(0, \frac{3}{2}, 1, 0, 0\right)$
7	(4, 6, 0, 0, −6)	No	42	Sí	$\left(3, \frac{5}{2}, 0, 0, 0\right)$
8	(0, 9, 4, −6, 0)	No	45	Sí	$\left(0, 0, \frac{5}{2}, \frac{9}{2}, 0\right)$

Propiedad de holgura complementaria: Dada la asociación entre variables que se proporciona en la tabla 6.7, las variables de la solución básica primal y de la solución básica dual complementaria satisfacen las relaciones de **holgura complementaria** que se presentan en la tabla 6.8. Aún más, esta relación es simétrica, de manera que las dos soluciones básicas son complementarias entre sí.

La razón por la que se usa el nombre de *holgura complementaria* para esta última propiedad es que enuncia (en parte) que para cada par de variables asociadas, si una de ellas tiene *holgura* en su restricción de no negatividad (una variable básica > 0), entonces la otra *no debe tener holgura* (variable no básica $= 0$). En la sección 6.2 se mencionó que esta propiedad tiene una interpretación económica útil en los problemas de programación lineal.

Ejemplo. Para ilustrar estas dos propiedades, considere de nuevo el problema de la Wyndor Glass Co., de la sección 3.1. En la tabla 6.9 se muestran sus ocho soluciones básicas (cinco factibles y tres no factibles). Entonces el problema dual (vea la tabla 6.1) también debe tener ocho soluciones básicas, cada una complementaria de una de las soluciones primales, como se muestra en la tabla 6.9.

Las tres soluciones BF que se obtuvieron por el método símplex para el problema primal son la primera, la quinta y la sexta soluciones que se muestran en la tabla 6.9. Ya se vio en la tabla 6.5 cómo se pueden leer las soluciones básicas complementarias del problema dual a partir del renglón 0: se comienza con los coeficientes de las variables de holgura y se sigue con las variables originales. Las otras soluciones básicas duales también se pueden identificar de la misma forma si se construye el renglón 0 para cada una de las otras soluciones básicas primales, para lo cual se utilizan las fórmulas que se ofrecen en la parte inferior de la tabla 5.8.

En forma alterna, para cada solución básica primal se puede usar la propiedad de holgura complementaria para identificar las variables básicas y no básicas de la solución básica dual complementaria, de manera que el sistema de ecuaciones que se presentó al principio de la sección se puede resolver de modo directo para obtener esta solución complementaria. Por ejemplo, considere la penúltima solución básica primal de la tabla 6.9, (4, 6, 0, 0, −6). Observe que x_1, x_2, y x_5 son

variables básicas, puesto que estas variables no son iguales a 0. En la tabla 6.7 se muestra que las variables duales asociadas son $(z_1 - c_1)$, $(z_2 - c_2)$ y y_3. En la tabla 6.8 se especifica que estas variables duales asociadas son variables no básicas de la solución básica complementaria

$$z_1 - c_1 = 0, \qquad z_2 - c_2 = 0, \qquad y_3 = 0.$$

En consecuencia, la forma aumentada de las restricciones funcionales en el problema dual,

$$\begin{aligned} y_1 \quad\;\; + 3y_3 - (z_1 - c_1) &= 3 \\ 2y_2 + 2y_3 - (z_2 - c_2) &= 5, \end{aligned}$$

se reduce a

$$\begin{aligned} y_1 \quad\;\; + 0 - 0 &= 3 \\ 2y_2 + 0 - 0 &= 5, \end{aligned}$$

de manera que $y_1 = 3$ y $y_2 = \frac{5}{2}$. Si se combinan estos valores con los valores de 0 de las variables no básicas se obtiene la solución básica $(3, \frac{5}{2}, 0, 0, 0)$, que se muestra en la columna de la derecha y el penúltimo renglón de la tabla 6.9. Observe que esta solución dual es factible para el problema dual porque las cinco variables satisfacen las restricciones de no negatividad.

Por último, observe que la tabla 6.9 muestra que $(0, \frac{3}{2}, 1, 0, 0)$ es la solución óptima para el problema dual, ya que es la solución básica *factible* con un valor mínimo de W (36).

Relaciones entre las soluciones básicas complementarias

A continuación se estudiarán las relaciones entre las soluciones básicas complementarias, comenzando por sus relaciones de *factibilidad*. Las columnas centrales de la tabla 6.9 proporcionan algunas ideas valiosas. Observe que para los pares de soluciones básicas complementarias, las respuestas de sí o no a la factibilidad también satisfacen una relación complementaria en la mayoría de los casos. En particular, con una sola excepción, siempre que una solución es factible, la otra no lo es. (También es posible que *ninguna* solución sea factible, como ocurre con el tercer par.) La única excepción es el sexto par, en el que se sabe que la solución del primal es óptima. La columna $Z = W$ sugiere la explicación. Como la sexta solución dual también es óptima (por la propiedad de soluciones óptimas complementarias), con $W = 36$, entonces las primeras cinco soluciones duales *no pueden ser factibles*, puesto que $W < 36$ (recuerde que el objetivo del problema dual es *minimizar W*). Por lo mismo, las dos últimas soluciones primales no pueden ser factibles porque $Z > 36$.

Esta explicación se apoya en la propiedad de dualidad fuerte que dice que las soluciones óptimas primal y dual tienen $Z = W$.

En seguida se establecerá la *extensión* de la propiedad de las soluciones óptimas complementarias de la sección 6.1 para las formas aumentadas de los dos problemas.

> **Propiedad de las soluciones básicas óptimas complementarias:** Cada solución básica *óptima* del *problema primal* tiene una **solución básica óptima complementaria** en el problema dual, donde los valores respectivos de las funciones objetivo (Z y W) son iguales. Dado el renglón 0 de la tabla símplex de la solución primal óptima, se puede encontrar la solución dual óptima complementaria $(\mathbf{y}^*, \mathbf{z}^* - \mathbf{c})$ como se muestra en la tabla 6.4.

Para revisar el razonamiento que fundamenta esta propiedad, observe que la solución dual $(\mathbf{y}^*, \mathbf{z}^* - \mathbf{c})$ debe ser factible para el problema dual, puesto que la condición de optimalidad para el problema primal requiere que *todas* las variables duales (inclusive las variables de superávit) sean *no negativas*. Como esta solución es *factible*, debe ser *óptima* para el problema dual debido a la propiedad de dualidad débil (como $W = Z$, entonces $\mathbf{y}^*\mathbf{b} = \mathbf{c}\mathbf{x}^*$ donde $\mathbf{x}^*$ es óptima para el problema primal).

Las soluciones básicas pueden ser clasificadas de acuerdo a si satisfacen o no cada una de las dos condiciones. Una es la *condición de factibilidad*, es decir, si todas las variables (inclusive las de holgura) de la solución aumentada son *no negativas*. La otra es la *condición de optimalidad*, es decir, si todos los coeficientes del renglón 0 (o sea, todas las variables de la solución básica complementaria) son *no negativos*. En la tabla 6.10 se resumen los términos que se emplean aquí para los diferentes tipos de soluciones básicas. Por ejemplo, en la tabla 6.9 las soluciones básicas

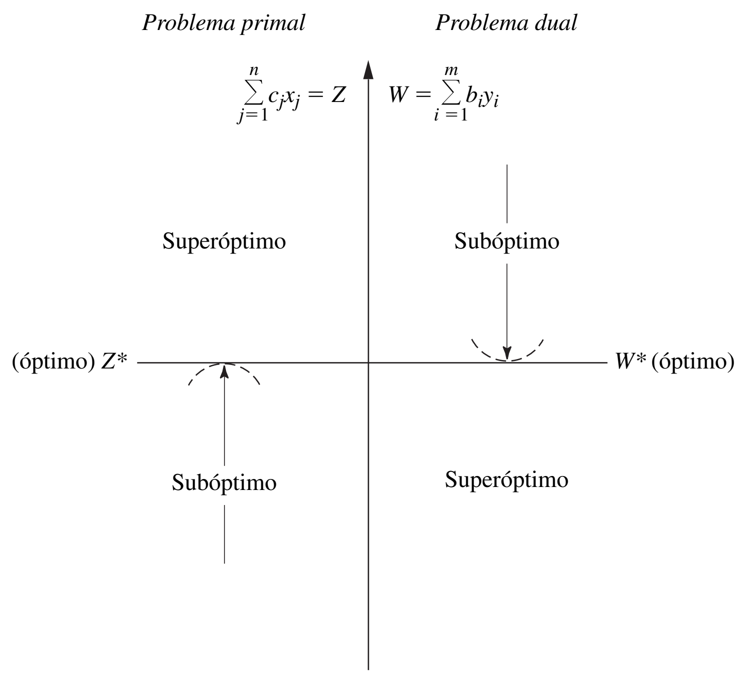


FIGURA 6.1
Intervalos de valores posibles de $Z = W$ para cierto tipo de soluciones básicas complementarias.

primales 1, 2, 4 y 5 son subóptimas, la 6 es óptima, la 7 y la 8 superóptimas y la 3 no es factible ni superóptima.

Dadas estas definiciones, en la tabla 6.11 se resumen las relaciones generales entre las soluciones básicas complementarias. En la figura 6.1 se muestra el intervalo de valores posibles (comunes) para las funciones objetivo ($Z = W$) que se obtiene para los tres pares de soluciones dados en la tabla 6.11 (el último par puede tener cualquier valor). Entonces, mientras que el método símplex trata en forma directa con soluciones básicas subóptimas y trabaja para lograr la optimalidad del problema primal, al mismo tiempo obtiene de manera indirecta las soluciones superóptimas complementarias y las mueve hacia la factibilidad del problema dual. A la inversa, algunas veces es

TABLA 6.10 Clasificación de las soluciones básicas

		¿Satisface la condición de optimalidad?	
		Sí	**No**
¿Es factible?	**Sí**	Óptima	Subóptima
	No	Superóptima	Ni factible ni superóptima

TABLA 6.11 Relaciones entre las soluciones básicas complementarias

Solución básica primal	Solución básica dual complementaria	Ambas soluciones básicas: ¿Primal factible?	Ambas soluciones básicas: ¿Dual factible?
Subóptima	Superóptima	Sí	No
Óptima	Óptima	Sí	Sí
Superóptima	Subóptima	No	Sí
Ni factible ni superóptima	Ni factible ni superóptima	No	No

más conveniente (o necesario) trabajar de modo directo con las soluciones básicas superóptimas y moverse hacia la factibilidad del problema primal, que es el propósito del método símplex dual descrito en la sección 7.1.

La tercera y cuarta columnas de la tabla 6.11 introducen otros dos términos comunes que se utilizan para describir un par de soluciones básicas complementarias. Se dice que las dos soluciones son **factibles primales** si la solución básica primal es factible, mientras que se llaman **factibles duales** si la solución básica dual complementaria es factible para el problema dual. Al usar esta terminología, el método maneja las soluciones factibles primales y se esfuerza por lograr también la factibilidad dual. Cuando lo logra, las dos soluciones básicas complementarias son óptimas para sus respectivos problemas.

Se ha comprobado que estas relaciones son útiles, en particular en el análisis de sensibilidad, como se verá más adelante en este capítulo.

6.4 ADAPTACIÓN A OTRAS FORMAS DEL PRIMAL

Hasta aquí se ha supuesto que el modelo del problema primal se encuentra en nuestra forma estándar. Sin embargo, al principio del capítulo se indicó que cualquier problema de programación lineal, ya sea que se encuentre en nuestra forma estándar o no, posee un problema dual. Por ello, esta sección está dedicada a estudiar la manera en que el problema dual cambia para otras formas del primal.

En la sección 4.6 se presentaron las formas diferentes a la estándar. Ahí se estableció que, si se desea, es posible convertir cada una en una forma estándar equivalente. Estas conversiones se resumen en la tabla 6.12. Con esto siempre se tiene la opción de convertir cualquier modelo en nuestra forma estándar y *después* construir su problema dual en la forma usual. Como ejemplo, en la tabla 6.13 se utiliza este procedimiento para el problema dual estándar (que también debe tener un dual). Observe que el resultado de todo lo que hicimos es justo el problema primal estándar. Como cualquier par de problemas primal y dual se pueden cambiar a estas formas, el dual del problema dual es siempre el problema primal. Por tanto, para cualquier problema primal y su problema dual, todas las relaciones entre ellos deben ser simétricas. Ésta es la propiedad de simetría que se estableció en la sección 6.1 (sin demostración), y que ahora en la tabla 6.13 se muestra por qué se cumple.

Una consecuencia de la propiedad de simetría es que todas las afirmaciones que se hicieron antes sobre las relaciones del problema dual con el problema primal también se cumplen en sentido inverso.

Otra consecuencia es que no importa a cuál de los problemas se le dé el nombre de primal y a cuál el de dual. En la práctica, se puede encontrar un problema de programación lineal que se ajuste a nuestra forma estándar y al que se le dé el nombre de problema dual. La convención es que el modelo que se formula para representar el problema real recibe el nombre de problema primal sin que importe qué forma tiene.

La ilustración de la forma de construir el problema dual para un problema primal no estándar no incluyó restricciones de igualdad ni variables no restringidas en signo. En realidad, para estas dos formas existe un camino más corto. Es posible demostrar (vea los problemas 6.4-7 y 6.4-2*a*)

TABLA 6.12 Conversiones de los modelos de programación lineal a la forma estándar

Forma no estándar	Forma estándar equivalente
Minimizar Z	Maximizar $(-Z)$
$\sum_{j=1}^{n} a_{ij}x_j \geq b_i$	$-\sum_{j=1}^{n} a_{ij}x_j \leq -b_i$
$\sum_{j=1}^{n} a_{ij}x_j = b_i$	$\sum_{j=1}^{n} a_{ij}x_j \leq b_i$ y $-\sum_{j=1}^{n} a_{ij}x_j \leq -b_i$
x_j sin restricción de signo	$x_j^+ - x_j^-$, $x_j^+ \geq 0$, $x_j^- \geq 0$

TABLA 6.13 Construcción del dual de un problema dual

Problema dual		*Convertido a forma estándar*
Minimizar $W = \mathbf{yb}$, sujeta a $\mathbf{yA} \geq \mathbf{c}$ y $\mathbf{y} \geq \mathbf{0}$.	$\longrightarrow$	Maximizar $(-W) = -\mathbf{yb}$, sujeta a $-\mathbf{yA} \leq -\mathbf{c}$ y $\mathbf{y} \geq \mathbf{0}$.
		$\downarrow$
Convertido a forma estándar		*Su problema dual*
Maximizar $Z = \mathbf{cx}$, sujeta a $\mathbf{Ax} \leq \mathbf{b}$ y $\mathbf{x} \geq \mathbf{0}$.	$\longleftarrow$	Minimizar $(-Z) = -\mathbf{cx}$, sujeta a $-\mathbf{Ax} \geq -\mathbf{b}$ y $\mathbf{x} \geq \mathbf{0}$.

que una *restricción de igualdad* en el primal debe manejarse igual que una restricción de la forma $\leq$ al construir el problema dual, excepto que debe *eliminarse* la restricción de no negatividad de la variable dual correspondiente (es decir, esta variable queda sin restricción de signo). Debido a la propiedad de simetría, la eliminación de una restricción de no negatividad en el problema primal afecta el problema dual sólo en que la restricción de desigualdad correspondiente cambia a una restricción de igualdad.

Otro atajo se refiere a las restricciones funcionales de la forma $\geq$ de un problema de maximización. El enfoque directo (pero no más largo) sería comenzar por convertir cada restricción de este tipo en la forma $\leq$

$$\sum_{j=1}^{n} a_{ij}x_j \geq b_i \longrightarrow -\sum_{j=1}^{n} a_{ij}x_j \leq -b_i.$$

Entonces, la construcción del problema dual en la forma usual dice que $-a_{ij}$ es el coeficiente de y_i en la restricción funcional j (que tiene la forma $\geq$) y $-b_i$ es el coeficiente de la función objetivo (que debe minimizarse), donde y_i también tiene restricción de no negatividad $y_i \geq 0$. Ahora suponga que se define una nueva variable $y_i' = -y_i$. Los cambios que ocurren al expresar el problema dual en términos de y_i' en lugar de y_i son: 1) los coeficientes de la variable se convierten en a_{ij} para la restricción funcional j y b_i para la función objetivo y 2) la restricción sobre la variable se convierte en $y_i' \leq 0$ (una *restricción de no positividad*). El atajo consiste en usar y_i' en lugar de y_i como variable dual para que los parámetros en la restricción original (a_{ij} y b_i) se conviertan de inmediato en los coeficientes de las variables en el problema dual.

A continuación se presenta una regla mnemotécnica para recordar cuáles deben ser las formas duales de las restricciones. En un problema de maximización, suele ser *común* que una restricción tenga la forma $\leq$ un *poco extraño* que esté en la forma $=$, y *bastante raro* que tenga la forma $\geq$. De manera similar, puede ser *común* que las restricciones en un problema de minimización tengan la forma $\geq$, un *poco extraño* que tenga la forma $=$, y *bastante raro* que tenga la forma $\leq$. En el caso de la restricción sobre una variable individual de cualquier tipo de problema, suele *ser común* que tenga restricción de no negatividad, un *poco extraño* que no tenga restricción (que la variable sea no restringida en signo) y *bastante raro* que esté restringida a valores *menores* o iguales a cero. Ahora recuerde la correspondencia entre los elementos de los problemas primal y dual que se muestran en la tabla 6.3; esto es, la restricción funcional i de un problema corresponde a la variable i en el otro y viceversa. El **método común-extraño-raro**, o **método CER**, señala que la forma de una restricción funcional o de la restricción sobre una variable del problema dual debe ser común, extraña o rara, lo que depende de que la forma del elemento correspondiente en el problema sea común, extraña o rara. En seguida se presenta un resumen.

El método CER para determinar la forma de las restricciones en el dual[3]

1. Formule el problema primal de cualquier forma, maximización o minimización, y el problema dual automáticamente quedará en la forma contraria.
2. Determine si cada forma de las restricciones funcionales y de las restricciones sobre las variables es *común*, *extraña* o *rara* de acuerdo con la tabla 6.14. La condición de las restricciones funcionales dependen de que el problema sea un problema de *maximización* (use la segunda columna) o un problema de *minimización* (use la tercera columna).
3. Por cada restricción sobre una *variable individual* en el problema dual, use la forma que tiene la misma condición que la restricción funcional en el problema primal que corresponde a esta variable dual (como se indica en la tabla 6.3).
4. Por cada *restricción funcional* en el problema dual, use la forma que tiene la misma condición que la restricción sobre la variable individual correspondiente del problema primal (como se indica en la tabla 6.3).

Las flechas entre la segunda y tercera columnas de la tabla 6.14 señalan la correspondencia entre las formas de las restricciones en el primal y el dual. Observe que la correspondencia siempre se establece entre una restricción funcional de un problema y una restricción sobre una variable individual del otro problema. Como el problema primal puede ser de maximización o minimización, por lo que el dual será entonces del tipo opuesto, la segunda columna de la tabla proporciona la forma para el que sea el problema de maximización y la tercera columna, la forma para el otro problema (el de minimización).

Para ilustrar lo que decimos, considere el ejemplo de terapia de radiación que se presentó al comienzo de la sección 3.4. Para mostrar la conversión en ambas direcciones en la tabla 6.14, se establece que la forma de maximización de este modelo es el problema primal, antes de usar la forma de minimización (original).

El problema primal en la forma de maximización se muestra en el lado izquierdo de la tabla 6.15. Se usa la columna izquierda de la tabla 6.14 para representar este problema y las flechas indican la forma del problema dual en la tercera columna. Estas mismas flechas se usan en la tabla 6.15 para mostrar el problema dual que se obtiene. (Debido a estas flechas las restricciones funcionales se colocaron al final del problema dual en lugar de su posición usual al principio.) Al lado de cada restricción en ambos problemas, se insertó (entre paréntesis) una C, E o R para determinar si la forma es común, extraña o rara. Como se prescribe en el método CER, la condición de cada restricción dual siempre es la misma que para la restricción correspondiente en el primal.

TABLA 6.14 Formas primal-dual correspondientes

Etiqueta	Problema primal (o problema dual)		Problema dual (o problema primal)
	Maximizar Z (o W)		Minimizar W (o Z)
	Restricción i:		Variable y_i (o x_i):
Común	$\leq$ forma	⟷	$y_i \geq 0$
Extraña	$=$ forma	⟷	No restringida
Rara	$\geq$ forma	⟷	$y_i' \leq 0$
	Variable x_j (o y_j):		Restricción j:
Común	$x_j \geq 0$	⟷	$\geq$ forma
Extraña	No restringida	⟷	$=$ forma
Rara	$x_j' \leq 0$	⟷	$\leq$ forma

[3] Esta regla mnemotécnica (y otra relacionada) sirve para recordar cuáles deben ser las formas de las restricciones duales sugeridas por Arthur T. Benjamin, un profesor de matemáticas de Harvey Mudd College. Un hecho interesante y extrañamente maravilloso sobre el profesor Benjamin es que es una de las más grandes calculadoras humanas del mundo que pueden realizar proezas como una multiplicación mental de números de seis dígitos. Usted podrá encontrar un análisis más profundo del método SOB así como su deducción en A.T. Benjamin: "Sensible Rules for Remembering Duals— The S-O-B Method", en *SIAM Review*, **37**(1): 85-87, 1995.

TABLA 6.15 Una forma primal-dual del ejemplo de terapia de radiación

	Problema primal		*Problema dual*	
	Maximizar $-Z = -0.4x_1 - 0.5x_2$,		Minimizar $W = 2.7y_1 + 6y_2 + 6y_3'$,	
	sujeta a		sujeta a	
(C)	$0.3x_1 + 0.1x_2 \leq 2.7$	⟷	$y_1 \geq 0$	(C)
(E)	$0.5x_1 + 0.5x_2 = 6$	⟷	y_2 no restringida en signo	(E)
(R)	$0.6x_1 + 0.4x_2 \geq 6$	⟷	$y_3' \leq 0$	(R)
	y		y	
(C)	$x_1 \geq 0$	⟷	$0.3y_1 + 0.5y_2 + 0.6y_3' \geq -0.4$	(C)
(C)	$x_2 \geq 0$	⟷	$0.1y_1 + 0.5y_2 + 0.4y_3' \geq -0.5$	(C)

Sin embargo, no había necesidad (más que con fines ilustrativos) de convertir el problema primal en la forma de maximización. Si se usa la forma de minimización original, el problema primal equivalente se muestra a la izquierda de la tabla 6.16. Ahora se usa la *tercera columna* de la tabla 6.14 para representar este problema, donde las flechas indican la forma del problema dual en la *segunda columna*. Estas mismas flechas en la tabla 6.16 muestran el problema dual que resulta en el lado derecho. De nuevo, las etiquetas de las restricciones ilustran la aplicación del método CER.

Los problemas primales de las tablas 6.15 y 6.16 son equivalentes y, de la misma manera, los dos problemas duales son también completamente equivalentes. La clave para reconocer esta equivalencia se encuentra en el hecho de que las variables en cada versión del problema dual son el negativo de aquellas en la otra versión ($y_1' = -y_1$, $y_2' = -y_2$, $y_3 = -y_3'$). Por tanto, en cada versión, si se usan las variables de la versión opuesta y si tanto la función objetivo como las restricciones se multiplican por -1, se obtiene la otra versión. (El problema 6.4-5 le pide que verifique esto.)

Si se desea ver **otro ejemplo** del uso del método CER para construir un problema dual, en la sección Worked Examples del sitio en internet de este texto se proporciona uno de ellos.

Si el método símplex se va a aplicar a cualquiera de los problemas, primal o dual que tenga variables restringidas a valores *no positivos* (por ejemplo, $y_3' \leq 0$ en el problema dual de la tabla 6.15), esta variable se puede sustituir por su contraparte no negativa (por ejemplo, $y_3 = -y_3'$).

Cuando se usan las variables artificiales para ayudar al método símplex a resolver un problema primal, la interpretación de dualidad del renglón 0 en la tabla símplex es la siguiente: como las variables artificiales tienen el papel de variables de holgura, sus coeficientes del renglón 0 ahora proporcionan los valores de las variables duales correspondientes a la solución básica complementaria para el problema dual. Como las variables artificiales se usan para sustituir el problema real con un problema artificial más conveniente, de hecho, este problema dual es el dual del problema artificial. Sin embargo, una vez que todas las variables artificiales se convierten en no básicas, se tienen de nuevo los problemas primal y dual reales. Con el método de las dos fases, deben retenerse las variables artificiales en la fase 2 para poder leer la solución dual completa en el renglón 0. Con el método de la gran M, como al iniciar se agregó M a los coeficientes de cada variable artificial en el renglón 0, el valor actual de cada variable dual correspondiente es el coeficiente actual de esta variable artificial *menos* M.

TABLA 6.16 Otra forma primal-dual del ejemplo de terapia de radiación

	Problema primal		*Problema dual*	
	Minimizar $Z = 0.4x_1 + 0.5x_2$,		Maximizar $W = 2.7y_1' + 6y_2' + 6y_3$,	
	sujeta a		sujeta a	
(R)	$0.3x_1 + 0.1x_2 \leq 2.7$	⟷	$y_1' \leq 0$	(R)
(E)	$0.5x_1 + 0.5x_2 = 6$	⟷	y_2' no restringida en signo	(E)
(C)	$0.6x_1 + 0.4x_2 \geq 6$	⟷	$y_3 \geq 0$	(C)
	y		y	
(C)	$x_1 \geq 0$	⟷	$0.3y_1' + 0.5y_2' + 0.6y_3 \leq 0.4$	(C)
(C)	$x_2 \geq 0$	⟷	$0.1y_1' + 0.5y_2' + 0.4y_3 \leq 0.6$	(C)

Por ejemplo, observe el renglón 0 de la tabla símplex final del ejemplo de terapia de radiación dado al final de la tabla 4.12. Después de restar M de los coeficientes de las variables artificiales $\bar{x}_4$ y $\bar{x}_6$, la solución óptima para el problema dual correspondiente de la tabla 6.15 está dada por los coeficientes de x_3, $\bar{x}_4$ y $\bar{x}_6$ como $(y_1, y_2, y_3') = (0.5, -1.1, 0)$. Como siempre, las variables de superávit de las dos restricciones funcionales corresponden a los coeficientes de x_1 y x_2 y son $z_1 - c_1 = 0$ y $z_2 - c_2 = 0$.

6.5 PAPEL DE LA TEORÍA DE LA DUALIDAD EN EL ANÁLISIS DE SENSIBILIDAD

Como se describe en ésta y las dos secciones siguientes, el análisis de sensibilidad consiste, en esencia, en la investigación del efecto que tiene sobre la solución óptima el hecho de hacer cambios en los valores de los parámetros del modelo a_{ij}, b_i y c_j. Sin embargo, al cambiar los valores de los parámetros en el problema primal se cambian también los valores correspondientes en el problema dual. Por tanto, se puede elegir qué problema se va a usar para investigar cada cambio. Gracias a las relaciones primal-dual que se presentaron en las secciones 6.1 y 6.3 (en especial la propiedad de soluciones básicas complementarias), es fácil ir de un problema a otro según se desee. En algunos casos es más conveniente analizar el problema dual en forma directa con objeto de determinar el efecto complementario sobre el problema primal. Se comenzará por considerar dos de estos casos.

Cambios en los coeficientes de una variable no básica

Suponga que los cambios que se hacen en el modelo original ocurren en los coeficientes de una variable que era no básica en la solución óptima original. ¿Cuál es el efecto de estos cambios sobre esta solución? ¿Todavía es factible? ¿Todavía es óptima?

Como la variable en cuestión es no básica (su valor es cero), el cambio en sus coeficientes no puede afectar la factibilidad de la solución, por lo cual, la pregunta que queda abierta en este caso es si todavía es óptima. Como se indica en las tablas 6.10 y 6.11, una pregunta equivalente es si la solución básica complementaria para el problema dual todavía es factible después de hacer estos cambios. Dado que los cambios afectan al problema dual nada más en una restricción, la pregunta se puede responder de manera sencilla al verificar si esta solución básica complementaria satisface la restricción revisada.

Este caso se ilustrará en la subsección correspondiente de la sección 6.7 después de desarrollar el ejemplo adecuado. En la sección Worked Examples del sitio en internet de este texto también se proporcionan otros ejemplos, tanto para este caso, como para el siguiente.

Introducción de una nueva variable

Como se indicó en la tabla 6.6, las variables de decisión del modelo suelen representar los niveles de las distintas actividades en consideración. En algunas situaciones, se seleccionan algunas de ellas de entre un grupo grande de actividades *posibles* en el que las restantes no fueron elegidas debido a que parecían ser menos atractivas. O quizás estas otras actividades no fueron visibles sino hasta después de formular y resolver el modelo original. De cualquier manera, la pregunta en este caso es si se justifica la inclusión de alguna de estas actividades no consideradas con anterioridad. En otras palabras, ¿cambiará la solución óptima original si se agrega cualquiera de estas actividades?

La inclusión de otra actividad equivale a introducir en el modelo una nueva variable, con los coeficientes apropiados en las restricciones funcionales y en la función objetivo. El único cambio que resulta en el problema dual es la introducción de una *nueva restricción* (vea la tabla 6.3).

Una vez hechos estos cambios, ¿la solución óptima original, junto con la nueva variable igual a cero (no básica), será todavía óptima para el problema primal? Igual que en el caso anterior, una pregunta equivalente es si la solución básica complementaria todavía es factible para el problema dual, e igual que antes, esta pregunta se puede contestar con sólo verificar si esta solución básica complementaria satisface una restricción, que en este caso se trata de la nueva restricción del problema dual.

Con propósitos de ilustración, suponga que en el problema de la Wyndor Glass Co., de la sección 3.1 se planea incluir en la línea de producción un tercer producto nuevo. Si x_{nueva} representa la tasa de producción de este artículo, el modelo revisado que resulta es:

$$\text{Maximizar} \qquad Z = 3x_1 + 5x_2 + 4x_{\text{nueva}},$$

sujeta a

$$\begin{aligned} x_1 \qquad\quad + 2x_{\text{nueva}} &\le 4 \\ 2x_2 + 3x_{\text{nueva}} &\le 12 \\ 3x_1 + 2x_2 + \ \ x_{\text{nueva}} &\le 18 \end{aligned}$$

y

$$x_1 \ge 0, \qquad x_2 \ge 0, \qquad x_{\text{nueva}} \ge 0.$$

Después de introducir las variables de holgura, la solución óptima original de este problema sin x_{nueva} (vea la tabla 4.8) era $(x_1, x_2, x_3, x_4, x_5) = (2, 6, 2, 0, 0)$. Si se incluye $x_{nueva} = 0$, ¿la solución todavía es óptima?

Para responder esta pregunta se verifica la solución básica complementaria para el problema dual. Como lo indica la *propiedad de soluciones básicas óptimas complementarias* en la sección 6.3, esta solución está dada en el renglón 0 de la tabla símplex *final* para el problema primal, con la localización que se muestra en la tabla 6.4 y se ilustra en la tabla 6.5. Por tanto, como se puede observar en el último renglón de la tabla 6.5 y el sexto renglón de la tabla 6.9, la solución es

$$(y_1, y_2, y_3, z_1 - c_1, z_2 - c_2) = \left(0, \frac{3}{2}, 1, 0, 0\right).$$

(De manera alterna, esta solución básica complementaria se puede obtener de la forma ilustrada en la sección 6.3 para la solución básica complementaria en el penúltimo renglón de la tabla 6.9.)

Dado que esta solución era óptima para el problema dual original, sin duda satisface las restricciones duales originales que se muestran en la tabla 6.1. Pero, ¿satisface la nueva restricción dual?

$$2y_1 + 3y_2 + y_3 \ge 4$$

Al sustituir esta solución se ve que

$$2(0) + 3\left(\frac{3}{2}\right) + (1) \ge 4$$

se satisface, por lo que esta solución dual aún es factible (es decir, óptima). En consecuencia, la solución primal original (2, 6, 2, 0, 0) junto con $x_{nueva} = 0$, todavía es óptima, y se concluye que *no* debe incluirse este nuevo producto en la producción.

Este enfoque también facilita la tarea de llevar a cabo un análisis de sensibilidad sobre los coeficientes de la nueva variable agregada al problema original. Con sólo verificar la nueva restricción dual se puede saber cuánto pueden cambiar los valores de estos parámetros antes de que afecten la factibilidad de la solución dual y la optimalidad de la solución primal.

Otras aplicaciones

Ya se han presentado otras dos aplicaciones importantes de la teoría de la dualidad al análisis de sensibilidad: los precios sombra y el *método símplex dual*. Como se vio en las secciones 4.7 y 6.2, la solución óptima dual $(y_1^*, y_2^*, \ldots, y_m^*)$ proporciona los precios sombra de los recursos respectivos que indican cuánto puede cambiar Z si se hacen cambios (pequeños) en las b_i (cantidades de recurso). El análisis resultante se ilustrará con cierto detalle en la sección 6.7.

En términos más generales, la interpretación económica del problema dual y el método símplex que se presentó en la sección 6.2 proporciona algunas ideas útiles para el análisis de sensibilidad.

Cuando se investigan los efectos de los cambios en las b_i o en las a_{ij} (de las variables básicas), la solución óptima original puede convertirse en una solución básica *superóptima* (vea la tabla 6.10). Si se desea *reoptimizar* para identificar la nueva solución óptima, se debe aplicar el método

símplex dual (que se presentó al final de las secciones 6.1 y 6.3) con inicio en esta solución básica. (Esta importante variante del método símplex se describirá en la sección 7.1.)

En la sección 6.1 se mencionó que en ocasiones es más eficiente resolver el problema dual mediante el método símplex con el fin de identificar una solución óptima para el problema primal. Cuando la solución se encuentra de esta manera, el análisis de sensibilidad del problema primal se lleva a cabo con la aplicación del procedimiento que se describe en las dos secciones siguientes directamente sobre el problema dual y después se infieren los efectos complementarios sobre el problema primal (por ejemplo, vea la tabla 6.11). Este enfoque sobre el análisis de sensibilidad es casi directo gracias a las estrechas relaciones primal-dual descritas en las secciones 6.1 y 6.3 (vea el problema 6.6-3.)

6.6 ESENCIA DEL ANÁLISIS DE SENSIBILIDAD

El trabajo del equipo de investigación de operaciones apenas comienza una vez que se aplica con éxito el método símplex para identificar una solución óptima para el modelo. Como se dijo al final de la sección 3.3, un supuesto de programación lineal es que todos los parámetros del modelo (a_{ij}, b_i, y c_j) son *constantes conocidas*. En realidad, los valores de los parámetros que se usan en el modelo casi siempre son sólo *estimaciones* basadas en una *predicción de las condiciones futuras*. Con frecuencia, los datos que se obtienen para desarrollar estas estimaciones son bastante burdos o no existen, así que los parámetros de la formulación original pueden representar tan sólo la opinión proporcionada por el personal de línea. Los datos pueden incluso representar estimaciones optimistas o pesimistas que protegen los intereses de los estimadores.

En consecuencia, un administrador razonable y el personal de investigación de operaciones deben mantener cierto escepticismo saludable respecto de los números originales que les proporciona la computadora y, en muchos casos, los tomarán nada más como un punto de partida para el análisis posterior del problema. Una solución "óptima" lo es sólo en lo que se refiere al modelo específico que se usa para representar el problema real, y esa solución no se convierte en una guía confiable para la acción hasta verificar que su comportamiento es bueno también para otras representaciones razonables del problema. Aún más, algunas veces los parámetros del modelo (en particular las b_i) se establecen como resultado de decisiones basadas en políticas administrativas (por ejemplo, la cantidad de ciertos recursos que se ponen a disposición de las actividades), y estas decisiones deben revisarse después de detectar sus consecuencias potenciales.

Por estas razones es importante llevar a cabo un **análisis de sensibilidad**, para investigar el efecto que tendría sobre la solución óptima que proporciona el método símplex el hecho de que los parámetros tomen otros valores posibles. En general, habrá algunos parámetros a los que se les pueda asignar cualquier valor razonable sin que afecten la optimalidad de esta solución. Sin embargo, también habrá parámetros con valores probables que lleven a una nueva solución óptima. Esta situación es seria, en particular si la solución original adquiere valores muy inferiores en la función objetivo, o tal vez no factibles.

Por tanto, un objetivo fundamental del análisis de sensibilidad es identificar los **parámetros sensibles** (es decir, los parámetros cuyos valores no pueden cambiar sin que cambie la solución óptima). Para coeficientes de la función objetivo que no están clasificados como sensibles, también puede resultar de gran utilidad determinar el *intervalo de valores* del parámetro para el que la solución óptima no cambia. (Este intervalo de valores se conoce como *intervalo permisible para ese coeficiente*.) En algunos casos, el cambio del valor de un parámetro en la columna del lado derecho dentro de una restricción funcional puede afectar la *factibilidad* de la solución BF óptima. Para manejar tales parámetros, es útil determinar el intervalo de valores para el que la solución BF óptima (con los valores ajustados de las variables básicas) seguirá siendo factible. (Este intervalo recibe el nombre de *intervalo permisible* por el lado derecho involucrado.) Este rango de valores es también el rango dentro del cual el *precio sombra* actual de la restricción correspondiente permanece válido. En la siguiente sección se describirán los procedimientos específicos para obtener este tipo de información.

Tal información es invaluable en dos sentidos. Primero, identifica los parámetros más importantes, por lo que se debe tener un cuidado especial para hacer estimaciones cercanas y seleccionar una solución que tenga un buen desempeño para la mayoría de los valores posibles. Segundo, identifica los parámetros que será necesario controlar muy de cerca cuando el estudio se implante.

Si se descubre que el valor real de un parámetro está fuera de su intervalo de valores permisibles, ésta es una señal incontrastable de que es necesario cambiar la solución.

Para problemas pequeños, la verificación del efecto de una variedad de cambios en los valores de los parámetros es directa con sólo aplicar de nuevo el método símplex para ver si cambia la solución óptima. Esto es en especial conveniente cuando se usa una formulación en hoja de cálculo. Una vez que el Solver se ha preparado para obtener una solución óptima, sólo se hacen los cambios deseados y se elige el botón de resolver otra vez.

Sin embargo, en problemas más grandes como los que se encuentran en la práctica, el análisis de sensibilidad requeriría de un esfuerzo computacional exorbitante si fuera necesario volver a aplicar el método símplex desde el principio para investigar cada cambio en el valor de un parámetro. Por fortuna, la idea fundamental que se presenta en la sección 5.3 casi elimina el esfuerzo computacional. En esencia, la idea fundamental revela *de inmediato* la forma en que los cambios al modelo original alterarían los números de la tabla símplex final (si se supone que se *duplica* la misma secuencia de operaciones algebraicas que realizó el método símplex la primera vez). Por tanto, después de hacer unos cuantos cálculos para actualizar esta tabla símplex, se puede verificar con facilidad si la solución BF óptima original ahora es no óptima (o no factible). Si es así, esta solución se usará como solución básica inicial para comenzar de nuevo el símplex (o el símplex dual) y encontrar una nueva solución óptima, si se desea. Si los cambios en el modelo no son muy grandes, sólo se requerirán unas cuantas iteraciones para obtener la nueva solución óptima a partir de esta solución básica inicial "avanzada".

Para describir este procedimiento con más detalle, considere la siguiente situación. Se ha empleado el método símplex para obtener una solución óptima para un modelo de programación lineal con valores específicos de los parámetros b_i, c_j y a_{ij}. Para iniciar el análisis de sensibilidad al menos uno de los parámetros será modificado. Después de hacer los cambios, sean $\bar{b}_i$, $\bar{c}_j$ y $\bar{a}_{ij}$ los valores de los distintos parámetros. Entonces, en notación matricial,

$$\mathbf{b} \rightarrow \bar{\mathbf{b}}, \qquad \mathbf{c} \rightarrow \bar{\mathbf{c}}, \qquad \mathbf{A} \rightarrow \bar{\mathbf{A}},$$

para el modelo revisado.

El primer paso es actualizar la tabla símplex final para que refleje estos cambios. En particular, se desea encontrar la tabla símplex final revisada que resultaría si se repitieran de manera *exacta* las mismas operaciones algebraicas (inclusive los mismos múltiplos de renglones sumados o restados a los otros renglones) que condujeron desde la tabla símplex inicial a la final. (Esto no es necesariamente lo mismo que reaplicar el método símplex, porque los cambios en la tabla inicial podrían ocasionar que el método símplex modifique algunas de las operaciones algebraicas utilizadas.) Si se utiliza la notación de la tabla 5.9, al igual que las fórmulas que la acompañan para la idea fundamental [(1) $\mathbf{t}^* = \mathbf{t} + \mathbf{y}^*\mathbf{T}$ y (2) $\mathbf{T}^* = \mathbf{S}^*\mathbf{T}$], se ve que la tabla símplex final revisada se calcula a partir de $\mathbf{y}^*$ y $\mathbf{S}^*$ (que no han cambiado) y la nueva tabla símplex inicial, como se muestra en la tabla 6.17. Observe que $\mathbf{y}^*$ y $\mathbf{S}^*$ juntas son los coeficientes de las *variables de holgura* de la tabla símplex final, donde el vector $\mathbf{y}^*$ (las variables duales) es igual a estos coeficientes del renglón 0 y la matriz $\mathbf{S}^*$ proporciona estos coeficientes en los otros renglones de la tabla. Entonces, sólo

TABLA 6.17 Tabla símplex final revisada que resulta de los cambios introducidos al modelo original

		Coeficiente de:			
	Ec.	Z	**Variables originales**	**Variables de holgura**	**Lado derecho**
Nueva tabla inicial	(0)	1	$-\bar{\mathbf{c}}$	$\mathbf{0}$	0
	(1, 2, . . . , m)	$\mathbf{0}$	$\bar{\mathbf{A}}$	$\mathbf{I}$	$\bar{\mathbf{b}}$
Tabla final revisada	(0)	1	$\mathbf{z}^* - \bar{\mathbf{c}} = \mathbf{y}^*\bar{\mathbf{A}} - \bar{\mathbf{c}}$	$\mathbf{y}^*$	$Z^* = \mathbf{y}^*\bar{\mathbf{b}}$
	(1, 2, . . . , m)	$\mathbf{0}$	$\mathbf{A}^* = \mathbf{S}^*\bar{\mathbf{A}}$	$\mathbf{S}^*$	$\mathbf{b}^* = \mathbf{S}^*\bar{\mathbf{b}}$

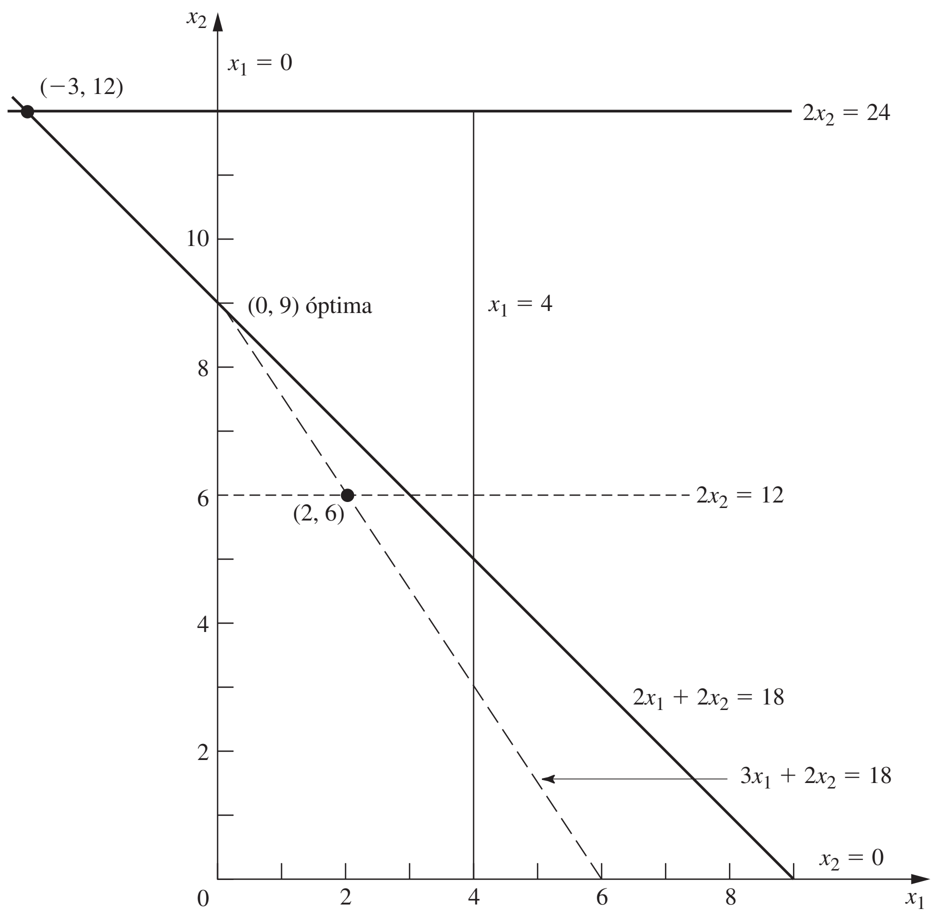
con usar **y***, **S***, y los números revisados en la tabla *inicial*, la tabla 6.17 revela la forma en que los números revisados en el resto de la tabla *final* se calculan de inmediato sin tener que repetir ninguna operación algebraica.

Ejemplo (variación 1 del modelo Wyndor). A manera de ilustración, suponga que se revisa el modelo original del problema de la Wyndor Glass Co., de la sección 3.1 según se muestra en la tabla 6.18.

Así, los cambios al modelo original son $c_1 = 3 \rightarrow 4$, $a_{31} = 3 \rightarrow 2$ y $b_2 = 12 \rightarrow 24$. La figura 6.2 muestra el efecto gráfico de estos cambios. En el caso del modelo original, el método símplex ya identificó la solución FEV óptima en el vértice (2, 6), que se encuentra en la intersección de las dos fronteras de restricción que se muestran como líneas punteadas $2x_2 = 12$ y $3x_1 + 2x_2 = 18$. Ahora, la revisión del modelo ha cambiado ambas fronteras por las que se muestran con líneas oscuras $2x_2 = 24$ y $2x_1 + 2x_2 = 18$. En consecuencia, la solución FEV anterior (2, 6) cambia ahora a la nueva intersección (−3, 12) que es una solución *no factible* en un vértice del modelo revisado. El procedimiento descrito en los párrafos anteriores encuentra este cambio de *manera algebraica* (en la forma aumentada). Más aún, lo hace de modo eficiente aunque se trate de problemas grandes donde el análisis gráfico es imposible.

Para realizar este procedimiento, se comienza por escribir los parámetros del modelo revisado en forma matricial:

$$\bar{\mathbf{c}} = [4,\ 5], \qquad \bar{\mathbf{A}} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 2 & 2 \end{bmatrix}, \qquad \bar{\mathbf{b}} = \begin{bmatrix} 4 \\ 24 \\ 18 \end{bmatrix}.$$



■ **FIGURA 6.2**
Cambio de la solución final en un vértice de (2, 6) a (−3, 12) de la revisión del problema de la Wyndor Glass Co., donde $c_1 = 3 \rightarrow 4$, $a_{31} = 3 \rightarrow 2$ y $b_2 = 12 \rightarrow 24$.

TABLA 6.18 El modelo original y el primer modelo revisado (variación 1) para realizar el análisis de sensibilidad con el modelo de la Wyndor Glass Co.

Modelo original	*Modelo revisado*
Maximizar $Z = [3, 5]\begin{bmatrix}x_1\\x_2\end{bmatrix}$, sujeta a $\begin{bmatrix}1 & 0\\0 & 2\\3 & 2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} \le \begin{bmatrix}4\\12\\18\end{bmatrix}$ y $\mathbf{x} \ge \mathbf{0}$.	Maximizar $Z = [4, 5]\begin{bmatrix}x_1\\x_2\end{bmatrix}$, sujeta a $\begin{bmatrix}1 & 0\\0 & 2\\2 & 2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} \le \begin{bmatrix}4\\24\\18\end{bmatrix}$ y $\mathbf{x} \ge \mathbf{0}$.

La nueva tabla símplex inicial que resulta se muestra en la parte superior de la tabla 6.19. Debajo se encuentra la tabla símplex final original (como se obtuvo en la tabla 4.8). Se dibujaron cuadros oscuros para marcar las partes de esta tabla símplex final que en definitiva *no cambian* junto con las modificaciones al modelo, esto es, los coeficientes de las variables de holgura tanto en el renglón 0 ($\mathbf{y}^*$) como en el resto de los renglones ($\mathbf{S}^*$). Así,

$$\mathbf{y}^* = [0, \tfrac{3}{2}, 1], \qquad \mathbf{S}^* = \begin{bmatrix}1 & \frac{1}{3} & -\frac{1}{3}\\ 0 & \frac{1}{2} & 0\\ 0 & -\frac{1}{3} & \frac{1}{3}\end{bmatrix}.$$

Es seguro que los coeficientes de las variables de holgura no cambian cuando se realizan las mismas operaciones algebraicas que originalmente realizó el método símplex, puesto que los coeficientes de estas mismas variables de la tabla símplex inicial no cambiaron.

TABLA 6.19 Obtención de la tabla símplex final revisada de la variación 1 del modelo de la Wyndor Glass Co.

	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	x_5	Lado derecho
Nueva tabla inicial	Z	(0)	1	−4	−5	0	0	0	0
	x_3	(1)	0	1	0	1	0	0	4
	x_4	(2)	0	0	2	0	1	0	24
	x_5	(3)	0	2	2	0	0	1	18
Tabla final del modelo original	Z	(0)	1	0	0	0	$\frac{3}{2}$	1	36
	x_3	(1)	0	0	0	1	$\frac{1}{3}$	$-\frac{1}{3}$	2
	x_2	(2)	0	0	1	0	$\frac{1}{2}$	0	6
	x_1	(3)	0	1	0	0	$-\frac{1}{3}$	$\frac{1}{3}$	2
Tabla final revisada	Z	(0)	1	−2	0	0	$\frac{3}{2}$	1	54
	x_3	(1)	0	$\frac{1}{3}$	0	1	$\frac{1}{3}$	$-\frac{1}{3}$	6
	x_2	(2)	0	0	1	0	$\frac{1}{2}$	0	12
	x_1	(3)	0	$\frac{2}{3}$	0	0	$-\frac{1}{3}$	$\frac{1}{3}$	−2

Sin embargo, como otras partes de la tabla símplex inicial cambiaron, habrá modificaciones en la tabla símplex final. Si se usan las fórmulas de la tabla 6.17, los números actualizados del resto de la tabla símplex final se calculan de la siguiente forma:

$$\mathbf{z}^* - \bar{\mathbf{c}} = [0, \tfrac{3}{2}, 1]\begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 2 & 2 \end{bmatrix} - [4, 5] = [-2, 0], \qquad Z^* = [0, \tfrac{3}{2}, 1]\begin{bmatrix} 4 \\ 24 \\ 18 \end{bmatrix} = 54,$$

$$\mathbf{A}^* = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & 1 \\ \frac{2}{3} & 0 \end{bmatrix},$$

$$\mathbf{b}^* = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 4 \\ 24 \\ 18 \end{bmatrix} = \begin{bmatrix} 6 \\ 12 \\ -2 \end{bmatrix}.$$

La tabla símplex final revisada que se obtiene se muestra en la parte inferior de la tabla 6.19.

En realidad, estos cálculos para obtener la tabla símplex final revisada se pueden simplificar de manera sustancial. Dado que ninguno de los coeficientes de x_2 cambió en el modelo original, ninguno de ellos puede cambiar en la tabla final, por lo que su cálculo se puede eliminar. Algunos otros parámetros originales (a_{11}, a_{21}, b_1, b_3) tampoco cambiaron; en consecuencia, otro atajo consiste en calcular sólo los *cambios incrementales* en la tabla símplex final en términos de los cambios incrementales de la tabla inicial y pasar por alto aquellos términos de la multiplicación de vectores o matrices que no tuvieron cambio en la tabla inicial. En particular, los únicos cambios incrementales que se produjeron en la tabla símplex inicial son $\Delta c_1 = 1$, $\Delta a_{31} = -1$ y $\Delta b_2 = 12$, por lo cual éstos son los únicos términos que deben considerarse. A continuación se muestra este enfoque de simplificación, donde se indica con un cero o un guión el lugar de los elementos no calculados.

$$\Delta(\mathbf{z}^* - \mathbf{c}) = \mathbf{y}^*\,\Delta\mathbf{A} - \Delta\mathbf{c} = [0, \tfrac{3}{2}, 1]\begin{bmatrix} 0 & — \\ 0 & — \\ -1 & — \end{bmatrix} - [1, —] = [-2, —].$$

$$\Delta Z^* = \mathbf{y}^*\,\Delta\mathbf{b} = [0, \tfrac{3}{2}, 1]\begin{bmatrix} 0 \\ 12 \\ 0 \end{bmatrix} = 18.$$

$$\Delta\mathbf{A}^* = \mathbf{S}^*\,\Delta\mathbf{A} = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 0 & — \\ 0 & — \\ -1 & — \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & — \\ 0 & — \\ -\frac{1}{3} & — \end{bmatrix}.$$

$$\Delta\mathbf{b}^* = \mathbf{S}^*\,\Delta\mathbf{b} = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 0 \\ 12 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \\ -4 \end{bmatrix}.$$

Al agregar estos incrementos a las cantidades originales en la tabla símplex final (parte media de la tabla 6.19) se obtiene la tabla símplex revisada final (parte inferior de la tabla 6.19).

Este *análisis incremental* proporciona también una idea general útil que indica que los cambios en la tabla símplex final deben ser *proporcionales* a cada cambio en la tabla símplex inicial. En la siguiente sección se ejemplificará la forma en que esta propiedad permite usar la interpolación o extrapolación lineal para determinar el intervalo de valores de un parámetro dado donde la solución básica final permanece tanto factible como óptima.

Después de obtener la tabla símplex final revisada, se convierte a la forma apropiada de eliminación de Gauss (si fuese necesario). En particular, la variable básica del renglón i debe tener un

coeficiente de 1 en ese renglón y 0 en todos los demás renglones (incluso en el renglón 0), para que la tabla esté en la forma apropiada para identificar y evaluar la solución básica actual. Entonces, si los cambios violan este requisito (lo que puede ocurrir sólo si se cambiaron los coeficientes de una variable básica original), se deberán realizar las operaciones necesarias para restablecer esta forma. Esta restauración se hace con el método de eliminación de Gauss, es decir, mediante aplicaciones sucesivas del paso 3 de una iteración del método símplex (vea el capítulo 4) como si cada variable que no cumple con el requisito fuera una variable entrante. Observe que este proceso puede causar otros cambios en la columna del *lado derecho*, por lo que la solución básica actual se podrá leer en esta columna cuando se haya recuperado por completo la forma apropiada de eliminación de Gauss.

En el caso del ejemplo, la tabla símplex final revisada que se muestra en la parte superior de la tabla 6.20 no está en la forma apropiada de eliminación gaussiana debido a la columna de la variable básica x_1. En particular, el coeficiente de x_1 en *su* renglón (3) es $\frac{2}{3}$ en lugar de 1, y tiene coeficientes distintos de cero (-2 y $\frac{1}{3}$) en los renglones 0 y 1. Para restablecer la forma apropiada se multiplica el renglón 3 por $\frac{3}{2}$; después, este nuevo renglón 3 multiplicado por 2 se suma al renglón 0; por último de este renglón 3 se resta del renglón 1. Las operaciones anteriores conducen a la forma apropiada de eliminación de Gauss que se muestra en la parte inferior de la tabla 6.20, que ahora se puede usar para identificar los nuevos valores de la solución básica actual (antes óptima):

$$(x_1, x_2, x_3, x_4, x_5) = (-3, 12, 7, 0, 0).$$

Como x_1 es negativa, esta solución básica ya no es factible, pero es *superóptima* (vea la tabla 6.10) y, por tanto, es *factible dual*, porque *todos* los coeficientes del renglón 0 son *no negativos*. Entonces, el método símplex dual puede ser útil para reoptimizar (si se desea), con un inicio en esta solución básica. (La rutina de análisis de sensibilidad del IOR Tutorial incluye esta opción.) Si se hace referencia a la figura 6.2 (y se pasan por alto las variables de holgura), el método símplex dual realiza sólo una iteración para moverse de la solución en el vértice $(-3, 12)$ a la solución FEV óptima en $(0, 9)$. (Con frecuencia, en el análisis de sensibilidad es útil identificar las soluciones que son óptimas para un conjunto de valores probables de los parámetros y después determinar cuál de estas soluciones es la que tiene de *manera permanente* el mejor desempeño ante los distintos valores posibles de los parámetros.)

Si esta solución básica $(-3, 12, 7, 0, 0)$ *no hubiera* sido factible en el primal ni en el dual (esto es, si la tabla hubiera tenido números negativos *tanto* en el *lado derecho* como en el renglón 0), se

TABLA 6.20 Conversión de la tabla símplex revisada final en la forma apropiada de Gauss para la variación 1 del modelo de la Wyndor Glass Co.

	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	x_5	Lado derecho
Tabla revisada final	Z	(0)	1	-2	0	0	$\frac{3}{2}$	1	54
	x_3	(1)	0	$\frac{1}{3}$	0	1	$\frac{1}{3}$	$-\frac{1}{3}$	6
	x_2	(2)	0	0	1	0	$\frac{1}{2}$	0	12
	x_1	(3)	0	$\frac{2}{3}$	0	0	$-\frac{1}{3}$	$\frac{1}{3}$	-2
Convertida a la forma apropiada	Z	(0)	1	0	0	0	$\frac{1}{2}$	2	48
	x_3	(1)	0	0	0	1	$\frac{1}{2}$	$-\frac{1}{2}$	7
	x_2	(2)	0	0	1	0	$\frac{1}{2}$	0	12
	x_1	(3)	0	1	0	0	$-\frac{1}{2}$	$\frac{1}{2}$	-3

hubieran podido introducir variables artificiales para convertirla en la forma apropiada para una tabla símplex inicial.[4]

Procedimiento general. Cuando se realizan pruebas para detectar el grado de *sensibilidad* de la solución óptima original a los distintos parámetros del modelo, el enfoque común es verificar cada parámetro (o al menos las c_j y b_i) en forma individual. Además de encontrar los intervalos permisibles descritos en la siguiente sección, esta verificación puede incluir cambios en la estimación inicial del valor del parámetro a otras posibilidades dentro del *intervalo de valores probables* (que incluyen los extremos del intervalo). Después, se pueden investigar algunas combinaciones de cambios simultáneos (como cambiar una restricción funcional completa). Cada vez que se hace un cambio en uno o más parámetros, debe aplicarse el procedimiento descrito e ilustrado. A continuación se resume este procedimiento.

Resumen del procedimiento para análisis de sensibilidad

1. *Revisión del modelo:* se hacen los cambios deseados en el modelo que se va a investigar.
2. *Revisión de la tabla símplex final:* Se emplea la idea fundamental (según el resumen de fórmulas al final de la tabla 6.17) para determinar los cambios que resultan en la tabla símplex final (vea un ejemplo en la tabla en 6.19).
3. *Conversión a la forma apropiada de eliminación de Gauss:* Se convierte esta tabla en la forma apropiada para identificar y evaluar la solución básica actual, para lo cual se aplica (según sea necesario) eliminación de Gauss (vea un ejemplo en la tabla 6.20).
4. *Prueba de factibilidad:* Se prueba la factibilidad de esta solución mediante la verificación de que todas las variables básicas de la columna del lado derecho aún tengan valores no negativos.
5. *Prueba de optimalidad:* Se verifica si esta solución es óptima (factible), mediante la comprobación de que todos los coeficientes de las variables no básicas del renglón 0 continúen no negativos.
6. *Reoptimización:* Si esta solución no pasa una de las pruebas, se puede obtener (si se desea) la nueva solución óptima a partir de la tabla actual como tabla símplex inicial (con las conversiones necesarias) por el método símplex o el símplex dual.

La rutina interactiva llamada *análisis de sensibilidad* incluida en el IOR Tutorial le permite practicar la aplicación de este procedimiento. Además, proporciona una demostración (también llamada *análisis de sensibilidad*) con **otro ejemplo**.

En el caso de problemas con sólo dos variables de decisión, el análisis gráfico proporciona una alternativa al procedimiento algebraico que se presentó con anterioridad para realizar análisis de sensibilidad. El IOR Tutorial incluye una rutina llamada *Graphical Method and Sensitivity Analysis* (Método gráfico y análisis de sensibilidad) para llevar a cabo este análisis gráfico de manera eficiente.

En la siguiente sección se presentará e ilustrará la aplicación de este procedimiento a cada una de las categorías más importantes de cambios al modelo original. Este estudio incluye, en parte, la extensión sobre el ejemplo que se introdujo en esta sección para investigar cambios en el modelo de la Wyndor Glass Co. De hecho, se comenzará por verificar en forma *individual* cada uno de los cambios anteriores. Al mismo tiempo, se integrarán al análisis de sensibilidad algunas aplicaciones de la teoría de la dualidad que se estudiaron en la sección 6.5.

6.7 APLICACIÓN DEL ANÁLISIS DE SENSIBILIDAD

El análisis de sensibilidad casi siempre comienza con la investigación de los cambios en los valores de las b_i, la cantidad del recurso i ($i = 1, 2, \ldots, m$) que se encuentra disponible para las actividades bajo consideración. La razón es que en general existe mayor flexibilidad cuando se establecen y ajustan estos valores que en el caso de los otros parámetros del modelo. Como ya se dijo en las secciones 4.7 y 6.2, la interpretación económica de las variables duales (las y_i) como precios sombra es muy útil para decidir cuáles son los cambios que deben estudiarse.

[4] Existe también un algoritmo primal-dual que se puede aplicar directamente a la tabla símplex sin conversión.

Caso 1: cambios en las b_i

Suponga que los únicos cambios al modelo actual consisten en el cambio de uno o más parámetros b_i ($i = 1, 2, \ldots, m$). En este caso, los *únicos* cambios que resultan en la tabla símplex final se encuentran en la columna del *lado derecho*. En consecuencia, la tabla símplex está en la forma apropiada de eliminación de Gauss y todos los coeficientes de las variables no básicas del renglón 0 aún son no negativos. En consecuencia, se pueden omitir los pasos de *conversión a la forma apropiada* de eliminación de Gauss y *prueba de optimalidad* del procedimiento general, después de revisar la columna del lado derecho, y la única pregunta que subsiste es si las variables básicas son no negativas (prueba de factibilidad).

Como se muestra en la tabla 6.17, cuando el vector de valores b_i se cambia de **b** a $\bar{\mathbf{b}}$, las fórmulas para calcular la nueva columna del *lado derecho* de la tabla símplex final son

Lado derecho del renglón 0 final: $Z^* = \mathbf{y}^*\bar{\mathbf{b}}$,
Lado derecho de los renglones $1, 2, \ldots, m$: $\mathbf{b}^* = \mathbf{S}^*\bar{\mathbf{b}}$.

(Vea en la parte inferior de la tabla 6.17 la localización del vector $\mathbf{y}^*$ y la matriz $\mathbf{S}^*$ que no cambiaron, en la tabla símplex final.) La primera ecuación tiene una interpretación económica natural que se relaciona con la interpretación económica de las variables duales que se presentaron al inicio de la sección 6.2. El vector $\mathbf{y}^*$ proporciona los valores óptimos de las variables duales, donde estos valores se interpretan como los *precios sombra* de los recursos respectivos. En particular, cuando $\mathbf{Z}^*$ representa el beneficio de utilizar la solución primal óptima $\mathbf{x}^*$ y cada b_i representa la cantidad de recurso i disponible, $\mathbf{y}_i^*$ indica qué tanto se puede incrementar la ganancia por unidad aumentada en b_i (para incrementos pequeños de b_i).

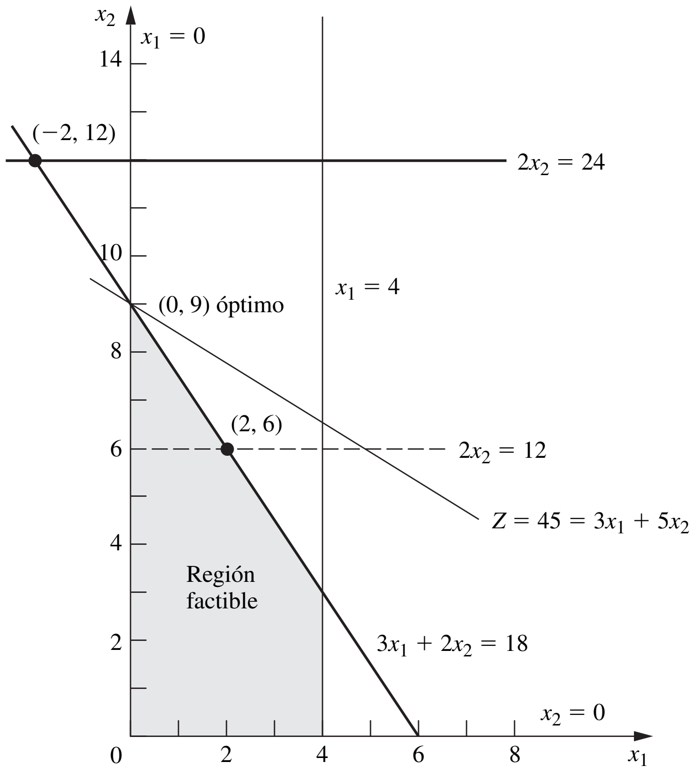
Ejemplo (variación 2 del modelo Wyndor). El análisis de sensibilidad del problema original de la Wyndor Glass Co., de la sección 3.1 comienza por examinar los valores óptimos de las variables duales y_i ($y_1^* = 0, y_2^* = \frac{3}{2}, y_3^* = 1$). Estos *precios sombra* indican el valor marginal de cada recurso i de las actividades (dos nuevos productos) en estudio, donde el valor marginal se expresa en las unidades de Z (miles de dólares de ganancia por semana). Como se dijo en la sección 4.7 (vea la figura 4.8), se puede aumentar la ganancia total debida a estas actividades en \$ 1 500 semanales (y_2^* multiplicado por \$ 1 000 por semana) por cada unidad adicional del recurso 2 (horas de producción a la semana en la planta 2) que quede disponible. Este aumento en la ganancia es válido en caso de cambios relativamente pequeños que no afecten la factibilidad de la solución básica actual (y por tanto no afectan los valores de y_i^*).

En consecuencia, el equipo de IO ha investigado la ganancia marginal posible debida a los otros usos actuales de este recurso para determinar si alguna es menor que \$ 1 500 semanales. Esta investigación puso de manifiesto que uno de los productos antiguos es mucho menos redituable. La tasa de producción de este producto ya se redujo a la cantidad mínima que justifica sus gastos de comercialización, pero se puede descontinuar, lo que proporcionaría 12 unidades adicionales del recurso 2 para los nuevos productos. En consecuencia, el siguiente paso es determinar la ganancia que se podría obtener de los nuevos productos si se hiciera esta transferencia. Lo anterior cambia b_2 de 12 a 24 en el modelo de programación lineal. En la figura 6.3 se muestra el efecto gráfico de este cambio, junto con el cambio de la solución en un vértice final de (2, 6) a (−2, 12). (Observe que esta figura es distinta de la 6.2 debido a la variación 1 del modelo Wyndor, ya que la restricción $3x_1 + 2x_2 \leq 18$ no cambió en este caso.)

De esta forma, en el caso de la variación 2 del modelo de la Wyndor, la única revisión al modelo original es el siguiente cambio en el vector de valores de las b_i:

$$\mathbf{b} = \begin{bmatrix} 4 \\ 12 \\ 18 \end{bmatrix} \longrightarrow \bar{\mathbf{b}} = \begin{bmatrix} 4 \\ 24 \\ 18 \end{bmatrix}.$$

entonces, sólo b_2 tiene nuevo valor.



■ **FIGURA 6.3**
Región factible de la variación 2 del modelo de la Wyndor Glass Co., donde $b_2 = 12 \rightarrow 24$.

Análisis de la variación 2. Cuando se aplica la idea fundamental (tabla 6.17), se encuentra que el efecto de este cambio en b_2 en la tabla símplex final original (parte media de la tabla 6.19) es que los elementos de la columna del lado derecho cambian a los siguientes valores:

$$Z^* = \mathbf{y}^*\overline{\mathbf{b}} = [0, \tfrac{3}{2}, 1]\begin{bmatrix} 4 \\ 24 \\ 18 \end{bmatrix} = 54,$$

$$\mathbf{b}^* = \mathbf{S}^*\overline{\mathbf{b}} = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 4 \\ 24 \\ 18 \end{bmatrix} = \begin{bmatrix} 6 \\ 12 \\ -2 \end{bmatrix}, \qquad \text{de manera que} \begin{bmatrix} x_3 \\ x_2 \\ x_1 \end{bmatrix} = \begin{bmatrix} 6 \\ 12 \\ -2 \end{bmatrix}.$$

De manera equivalente, como el único cambio en el modelo original es $\Delta b_2 = 24 - 12 = 12$, se puede usar el análisis incremental para calcular estos mismos valores con mayor rapidez. El análisis incremental involucra el cálculo de los *incrementos* en los valores de la tabla causados por el cambio (o cambios) en el modelo original, y después la suma de estos incrementos a los valores originales. En este caso, los incrementos en Z^* y $\mathbf{b}^*$ son

$$\Delta Z^* = \mathbf{y}^*\Delta\mathbf{b} = \mathbf{y}^*\begin{bmatrix} \Delta b_1 \\ \Delta b_2 \\ \Delta b_3 \end{bmatrix} = \mathbf{y}^*\begin{bmatrix} 0 \\ 12 \\ 0 \end{bmatrix},$$

$$\Delta\mathbf{b}^* = \mathbf{S}^*\,\Delta\mathbf{b} = \mathbf{S}^*\begin{bmatrix} \Delta b_1 \\ \Delta b_2 \\ \Delta b_3 \end{bmatrix} = \mathbf{S}^*\begin{bmatrix} 0 \\ 12 \\ 0 \end{bmatrix}.$$

Por tanto, si se usa el segundo componente de $\mathbf{y}^*$ y la segunda columna de $\mathbf{S}^*$, los únicos cálculos necesarios son

Recuadro de aplicación

Pacific Lumber Company (PALCO) es una compañía maderera de gran tamaño cuya casa matriz se encuentra en Scotia, California. La empresa cuenta con más de 200 000 acres de terrenos forestales altamente productivos que soportan cinco molinos ubicados en Humboldt County, al norte de California. Los terrenos incluyen algunas de las arboledas de secuoyas más espectaculares del mundo que han sido dadas o vendidas a bajo costo para preservarse como parques. PALCO administra los terrenos restantes de manera intensa para la producción sustentable de madera, sujeta a las prácticas legales más estrictas en cuanto a bosques. Debido a que los bosques de PALCO son casa de muchas especies silvestres, entre ellas especies en riesgo de extinción como, por ejemplo, lechuzas con lunares y mérgulos marmoleados, deben cumplir las disposiciones del Acta Federal de Especies en Peligro de Extinción.

Con el fin de obtener un plan de cosecha sostenida en todos los terrenos, la gerencia de PALCO contrató a un equipo de consultores en investigación de operaciones para desarrollar un plan a 120 años, de 12 periodos y a largo plazo para la administración del ecosistema forestal. Dicho equipo de consultores realizó esta tarea luego de plantear y aplicar un modelo de programación lineal para optimizar todas las operaciones y ganancias de la madera de la compañía después de satisfacer ciertas restricciones. El modelo fue enorme y contaba con 8 500 restricciones funcionales y 353 000 variables de decisión.

La gran cantidad de incógnitas que se tuvo que estimar para determinar cuáles deberían ser los parámetros del modelo constituyó un gran reto cuando llegó la etapa de aplicarlo. Los factores principales que provocaron dichas incógnitas fueron las continuas fluctuaciones de la oferta y la demanda del mercado, los costos de los troncos de madera y las regulaciones ambientales. Por tanto, el equipo de investigación de operaciones hizo un uso intensivo del *análisis detallado de sensibilidad*. El plan de cosecha sostenida resultante *incrementó la utilidad presente neta* de la compañía en *más* de **398 millones de dólares** a la vez que generó una mejor mezcla de superficie para el hábitat de los animales.

Fuente: L.R. Fletcher, H. Alden, S. P. Holmen, D. P. Angelis y M. J. Etzehouser. "Long-Termi Forest Ecosystem Planning at Pacific Lumber", en *Interfaces*, **29**(1): 90-112, enero-febrero de 1999. (Nuestro sitio en internet www.mhhe.com/hillier cuenta con una liga a este artículo.)

$$\Delta Z^* = \frac{3}{2}(12) = 18, \qquad \text{así } Z^* = 36 + 18 = 54$$

$$\Delta b_1^* = \frac{1}{3}(12) = 4, \qquad \text{así } b_1^* = 2 + 4 = 6,$$

$$\Delta b_2^* = \frac{1}{2}(12) = 6, \qquad \text{así } b_2^* = 6 + 6 = 12,$$

$$\Delta b_3^* = -\frac{1}{3}(12) = -4, \qquad \text{así } b_3^* = 2 - 4 = -2,$$

donde los valores originales de estas cantidades son obtenidos de la columna del lado derecho de la tabla símplex final original (parte media de la tabla 6.19). La tabla símplex final revisada corresponde por completo a la tabla símplex final original, excepto por la columna del lado derecho que tiene estos nuevos valores.

Por tanto, la solución básica actual (antes óptima) se ha convertido en

$$(x_1, x_2, x_3, x_4, x_5) = (-2, 12, 6, 0, 0),$$

TABLA 6.21 Datos de la variación 2 del modelo de la Wyndor Glass Co.

Parámetros del modelo

$c_1 = 3,$	$c_2 = 5$	$(n = 2)$
$a_{11} = 1,$	$a_{12} = 0,$	$b_1 = 4$
$a_{21} = 0,$	$a_{22} = 2,$	$b_2 = 24$
$a_{31} = 3,$	$a_{32} = 2,$	$b_3 = 18$

Tabla símplex final después de la reoptimización

Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	x_5	Lado derecho
Z	(0)	1	$\frac{9}{2}$	0	0	0	$\frac{5}{2}$	45
x_3	(1)	0	1	0	1	0	0	4
x_2	(2)	0	$\frac{3}{2}$	1	0	0	$\frac{1}{2}$	9
x_4	(3)	0	−3	0	0	1	−1	6

que no pasa la prueba de optimalidad porque tiene un valor negativo. Ahora se puede aplicar el método símplex dual, a partir de esta tabla símplex revisada, para encontrar la nueva solución óptima. Este método conduce, en una sola iteración, a la nueva tabla símplex final que se muestra en la tabla 6.21. (En forma alterna, se pudo haber aplicado el método símplex desde el principio y, en este caso, también se hubiera llegado a esta tabla final en una sola iteración.) Esta tabla símplex indica que la nueva solución óptima es

$$(x_1, x_2, x_3, x_4, x_5) = (0, 9, 4, 6, 0),$$

con $Z = 45$, lo que proporciona un incremento en la utilidad de 9 unidades ($9 000 semanales) sobre el valor anterior $Z = 36$. El hecho de que $x_4 = 6$ indica que 6 de las 12 unidades adicionales del recurso 2 no se usan con esta solución.

Con base en estos resultados, si $b_2 = 24$, el producto antiguo relativamente no redituable se descontinuará y las 6 unidades del recurso 2 quedarán para algún uso futuro. Como y_3^* aún es positiva, se realiza un estudio similar sobre la posibilidad de cambiar la asignación del recurso 3, pero la decisión resultante es conservar la asignación actual. En consecuencia, el modelo de programación lineal actual en este punto (variación 2) tiene los valores de los parámetros y la solución óptima que se muestran en la tabla 6.21. Este modelo se usará como punto de partida para investigar otros tipos de cambios en el modelo, más adelante en esta sección. Sin embargo, antes de estudiar otros casos, se ofrece un panorama más amplio del caso actual.

Intervalo permisible para un lado derecho. Aunque $\Delta b_2 = 12$ resultó ser un incremento de b_2 demasiado grande para mantener la factibilidad (y por ende la optimalidad) con la solución básica donde x_1, x_2 y x_3 son las variables básicas (parte media de la tabla 6.19), el análisis incremental anterior muestra cuán grande puede ser el incremento para mantener su carácter de factible. En particular, observe que

$$b_1^* = 2 + \frac{1}{3}\Delta b_2,$$
$$b_2^* = 6 + \frac{1}{2}\Delta b_2,$$
$$b_3^* = 2 - \frac{1}{3}\Delta b_2,$$

donde estas tres cantidades son los valores respectivos x_3, x_2 y x_1, para esta solución básica. La solución es factible y, por tanto, óptima, siempre y cuando las tres cantidades continúen siendo no negativas.

$$2 + \frac{1}{3}\Delta b_2 \ge 0 \quad \Rightarrow \quad \frac{1}{3}\Delta b_2 \ge -2 \quad \Rightarrow \quad \Delta b_2 \ge -6,$$
$$6 + \frac{1}{2}\Delta b_2 \ge 0 \quad \Rightarrow \quad \frac{1}{2}\Delta b_2 \ge -6 \quad \Rightarrow \quad \Delta b_2 \ge -12,$$
$$2 - \frac{1}{3}\Delta b_2 \ge 0 \quad \Rightarrow \quad 2 \ge \frac{1}{3}\Delta b_2 \quad \Rightarrow \quad \Delta b_2 \le 6.$$

Por tanto, como $b_2 = 12 + \Delta b_2$, la solución permanecerá factible sólo si

$$-6 \le \Delta b_2 \le 6, \qquad \text{es decir,} \qquad 6 \le b_2 \le 18.$$

(Verifique esto en la gráfica de la figura 6.3.) Como se introdujo en la sección 4.7, este intervalo de valores de b_2 se conoce como su *intervalo permisible* para permanecer factible.

> Para cualquier b_i, su **intervalo permisible** para permanecer factible es el intervalo de valores en el que la solución BF óptima[5] (con los valores ajustados para las variables bási-

[5] Cuando existe más de una solución BF óptima para el modelo actual (antes de cambiar b_i), aquí se hace referencia a la que se obtuvo con el método símplex.

cas) es factible. Así, el *precio sombra* de b_i aún es válido para evaluar el efecto del cambio en b_i sobre Z sólo si b_i sigue dentro de este intervalo permisible. (Se supone que el cambio en el valor de esta b_i es el único cambio en el modelo.) Los valores ajustados de las variables básicas se obtienen a partir de la fórmula $\mathbf{b}^* = \mathbf{S}^*\overline{\mathbf{b}}$. El cálculo del intervalo permisible para permanecer factibles está basado en el hecho de encontrar el intervalo de valores de b_i tales que $\mathbf{b}^* \geq \mathbf{0}$.

Muchos paquetes de software de programación lineal utilizan esta técnica para generar de manera automática el intervalo permisible para permanecer factible para cada b_i. (También se usa una técnica similar, que se presentó en los casos 2*a* y 3, para generar un intervalo de *valores permisibles* para permanecer factibles para cada c_j.) En el capítulo 4 se mostraron las salidas respectivas de Excel Solver y LINDO en las figuras 4.10 y A4.2. La tabla 6.22 resume estos mismos resultados respecto de b_i del modelo original de la Wyndor. Por ejemplo, tanto el *incremento* como el *decremento permisibles* de b_2 son 6, es decir, $-6 \leq \Delta b_2 \leq 6$. El análisis en el párrafo anterior muestra cómo se calculan estas dos cantidades.

Análisis de cambios simultáneos en los lados derechos. Cuando cambian varios valores b_i al mismo tiempo, de nuevo se puede usar la fórmula $\mathbf{b}^* = \mathbf{S}^*\overline{\mathbf{b}}$ para ver cómo cambian los lados derechos de la tabla final. Si todos los lados derechos son todavía no negativos, la prueba de factibilidad indica que la solución revisada obtenida en la tabla aún es factible. Como el renglón 0 no cambió, que sea factible implica que esta solución también es óptima.

Aunque este enfoque funciona para verificar el efecto de un conjunto *específico* de cambios en las b_i, no da un panorama amplio de cuánto pueden cambiar de manera simultánea a partir de sus valores originales antes de que la solución revisada sea no factible. Con frecuencia, como parte del análisis postóptimo, la administración de las organizaciones se interesa en investigar el efecto de distintos cambios en las políticas de decisión (como las cantidades de recursos que se ponen a disposición de las actividades que se estudian) que determinan los lados derechos. En lugar de considerar sólo un conjunto específico de cambios, quizá la administración desee estudiar las *direcciones* de éstos donde algunos lados derechos aumentan y otros diminuyen. Los precios sombra son invaluables para este tipo de estudios. Sin embargo, sólo son válidos para evaluar el efecto de los cambios en Z sólo dentro de ciertos intervalos de cambios. Para cada b_i, el *intervalo permisible* para permanecer factible proporciona los valores de este intervalo si *ninguna* otra b_i cambia al mismo tiempo. ¿En qué se convierten estos *intervalos permisibles* cuando algunas b_i tienen cambios simultáneos?

Una respuesta parcial a esta pregunta la proporciona la regla de 100%; esta regla combina los *cambios permisibles* (aumento o disminución) de las b_i que se proporcionan en las dos últimas columnas de la tabla 6.22.

Regla de 100% para cambios simultáneos en los lados derechos: Los precios sombra permanecen válidos para predecir el efecto de cambiar al mismo tiempo los lados derechos de algunas restricciones funcionales cuando los cambios no son muy grandes. Para verificar si son suficientemente pequeños, para cada cambio se calcula el porcentaje del cambio permitido (aumento o disminución) para que ese lado derecho siga dentro de su intervalo permisible y factible. Si la *suma* de los porcentajes de cambios *no* excede 100%, es definitivo que los precios sombra todavía serán válidos. (Si la suma excede 100%, entonces no hay certeza.)

TABLA 6.22 Salida típica de un software para análisis de sensibilidad de los lados derechos del modelo de la Wyndor Glass Co.

Restricción	Precio sombra	LD actual	Incremento permisible	Decremento permisible
Planta 1	0	4	∞	2
Planta 2	1.5	12	6	6
Planta 3	1	18	6	6

Ejemplo (variación 3 del modelo de la Wyndor). Para ilustrar esta regla, considere la *variación* 3 del modelo de la Wyndor Glass Co., que expone el modelo original con los siguientes cambios en el vector del lado derecho:

$$\mathbf{b} = \begin{bmatrix} 4 \\ 12 \\ 18 \end{bmatrix} \rightarrow \overline{\mathbf{b}} = \begin{bmatrix} 4 \\ 15 \\ 15 \end{bmatrix}.$$

Los cálculos para la regla de 100% en este caso son

$$b_2\text{: } 12 \rightarrow 15. \quad \text{Porcentaje de incremento permisible} = 100\left(\frac{15-12}{6}\right) = 50\%$$

$$b_3\text{: } 18 \rightarrow 15. \quad \text{Porcentaje de decremento permisible} = 100\left(\frac{18-15}{6}\right) = \underline{50\%}$$

$$\text{Suma} = 100\%$$

Como el resultado de la suma de 100% es justo la cantidad para no exceder de 100%, los precios sombra son definitivamente válidos para predecir el efecto de estos cambios en Z. En particular, como los precios sombra respectivos de b_2 y b_3 son 1.5 y 1, el cambio que resulta en Z sería

$$\Delta Z = 1.5(3) + 1(-3) = 1.5,$$

con base en lo cual Z^* se incrementará de 36 a 37.5.

En la figura 6.4 se muestra la región factible de este modelo revisado (las líneas punteadas son las localizaciones originales de las fronteras de restricción revisadas). La solución óptima ahora es la solución FEV (0, 7.5), que da

$$Z = 3x_1 + 5x_2 = 0 + 5(7.5) = 37.5,$$

exactamente como lo pronosticaron los precios sombra. Sin embargo, observe lo que pasaría si se aumentara b_2 hasta más de 15 o se redujera b_3 a menos de 15, de modo que la suma de los porcentajes de los cambios permisibles excediera de 100%. Este exceso causaría que la solución en un vértice, antes óptima, se deslice hacia la izquierda del eje x_2 ($x_1 < 0$), por lo que esta solución *no factible* ya no es óptima. En consecuencia, los precios sombra anteriores ya no son válidos para pronosticar el nuevo valor de Z^*.

Caso 2*a*: cambios en los coeficientes de una variable no básica

Considere una variable específica x_j (j fija) que sea no básica en la solución óptima dada en la tabla símplex final. En el caso 2*a*, el único cambio del modelo actual es que uno o más coeficientes de esta variable —$c_j, a_{1j}, a_{2j}, \ldots, a_{mj}$— cambiaron. Entonces, si $\overline{c}_j$ y $\overline{a}_{ij}$ denotan los nuevos valores de estos parámetros con $\overline{\mathbf{A}}_j$ (columna j de la matriz $\overline{\mathbf{A}}$) como el vector que contiene a_{ij}, se tiene

$$c_j \longrightarrow \overline{c}_j, \qquad \mathbf{A}_j \longrightarrow \overline{\mathbf{A}}_j$$

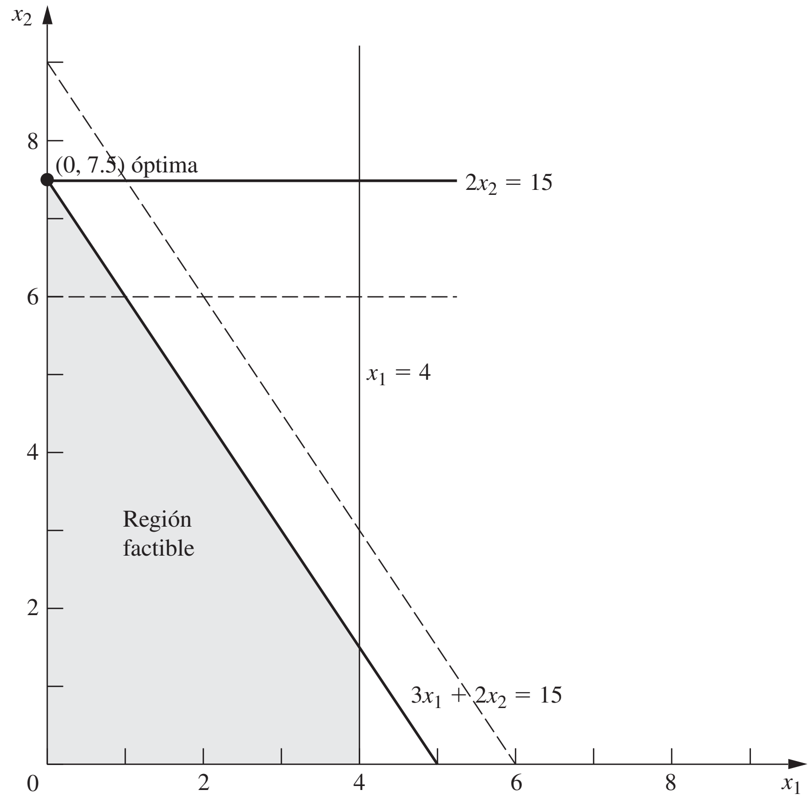
para el modelo revisado.

Como se describió al principio de la sección 6.5, la teoría de dualidad proporciona una manera muy conveniente de verificar estos cambios. En particular, si la solución básica *complementaria* $\mathbf{y}^*$ en el problema dual todavía satisface la restricción dual que cambió, entonces la solución óptima original del problema primal *permanece óptima* como tal. Por el contrario, si $\mathbf{y}^*$ viola esta restricción dual, entonces esta solución primal ya *no es óptima*.

Si la solución óptima cambió y se desea encontrar la nueva, se puede hacer en forma bastante sencilla. Sólo debe aplicarse la idea fundamental a la columna de x_j revisada (la única que cambia) en la tabla símplex final. En particular, las fórmulas de la tabla 6.17 se reducen de la siguiente manera:

Coeficiente de x_j del renglón 0 final: $\quad z_j^* - \overline{c}_j = \mathbf{y}^*\overline{\mathbf{A}}_j - \overline{c}_j,$

Coeficiente de x_j de los renglones 1 a m finales: $\quad \mathbf{A}_j^* = \mathbf{S}^*\overline{\mathbf{A}}_j.$



■ **FIGURA 6.4**
Región factible de la variación 3 del modelo de la Wyndor Glass Co., donde $b_2 = 12 \rightarrow 15$ y $b_3 = 18 \rightarrow 15$.

Debido a que la solución básica actual ya no es óptima, el nuevo valor de $z_j^* - c_j$ será ahora el que tiene coeficiente negativo en el renglón 0; así, se inicia el método símplex con x_j como variable básica entrante inicial.

Observe que este procedimiento es una versión simplificada del procedimiento general que se resumió al final de la sección 6.6. Los pasos 3 y 4 (conversión a la forma apropiada de eliminación gaussiana y prueba de factibilidad) se eliminaron por no ser relevantes, ya que la única columna que cambia en la revisión de la tabla símplex final (antes de reoptimizar) es la de la variable no básica x_j. El paso 5 (prueba de optimalidad) se sustituyó por una prueba más rápida que debe realizarse después del paso 1 (revisión del modelo). Sólo cuando esta prueba revele que la solución óptima cambia, y se desee encontrar la nueva solución, tendrán que aplicarse los pasos 2 y 6 (revisión de la tabla símplex final y reoptimización).

Ejemplo (variación 4 del modelo de la Wyndor). En razón de que x_1 es no básica en la solución óptima actual (vea la tabla 6.21) para la variación 2 del problema de la Wyndor Glass Co., el siguiente paso del análisis de sensibilidad es comprobar si cualquier cambio razonable en la estimación de los coeficientes de x_1 puede sugerir que se introduzca el producto 1. El conjunto de cambios que pueden ser realistas para hacer el producto 1 más atractivo sería restablecer $c_1 = 4$ y $a_{31} = 2$. En lugar de explorar cada uno de estos cambios en forma independiente (como se hace con frecuencia en el análisis de sensibilidad), se considerarán juntos. Entonces, los cambios bajo consideración son

$$c_1 = 3 \longrightarrow \bar{c}_1 = 4, \qquad \mathbf{A}_1 = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} \longrightarrow \overline{\mathbf{A}}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}.$$

Estos dos cambios en la variación 2 proporcionan la *variación* 4 del modelo de la Wyndor. En realidad, la variación 4 es equivalente a la variación 1 considerada en la sección 6.6 y descrita en

la figura 6.2, ya que la variación 1 combina estos dos cambios con el cambio en el modelo original ($b_2 = 12 \rightarrow 24$) que da la variación 2. Sin embargo, la diferencia clave respecto del tratamiento de la variación 1 en la sección 6.6 es que el análisis de la variación 4 trata a la variación 2 como si fuera el modelo original; de esta forma, el punto de partida es la tabla símplex dada en la tabla 6.21 donde ahora x_1 es una variable no básica.

El cambio en a_{31} hace que la región factible cambie de la que se muestra en la figura 6.3 a la región correspondiente de la figura 6.5. El cambio en c_1 hace que la función objetivo $Z = 3x_1 + 5x_2$ cambie a $Z = 4x_1 + 5x_2$. La figura 6.5 muestra la recta de la función objetivo $Z = 45 = 4x_1 + 5x_2$ que pasa por la solución óptima actual (0, 9), por lo que este punto aún es óptimo después de los cambios en a_{31} y c_1.

Para emplear la teoría de la dualidad y llegar a esta misma conclusión, observe que los cambios en c_1 y a_{31} llevan a una sola restricción revisada para el problema dual, que es la restricción $a_{11}y_1 + a_{21}y_2 + a_{31}y_3 \geq c_1$. Tanto esta restricción revisada como la $\mathbf{y}^*$ actual (coeficientes de las variables de holgura en el renglón 0 de la tabla 6.21) son

$$y_1^* = 0, \qquad y_2^* = 0, \qquad y_3^* = \frac{5}{2},$$

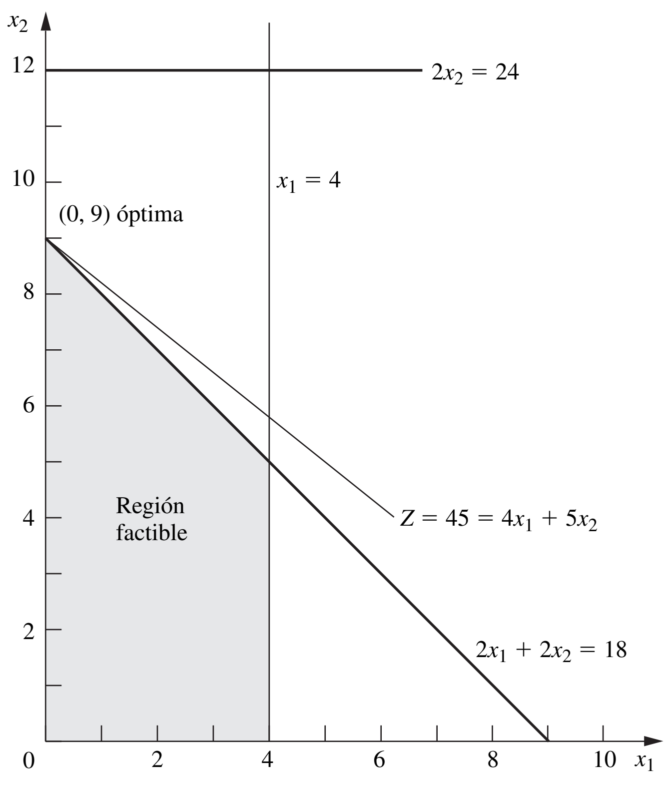
$$y_1 + 3y_3 \geq 3 \longrightarrow y_1 + 2y_3 \geq 4,$$

$$0 + 2\left(\frac{5}{2}\right) \geq 4.$$

Como $\mathbf{y}^*$ *todavía* satisface la restricción revisada, la solución primal actual (tabla 6.21) todavía es óptima.

Dado que esta solución todavía es óptima, no existe la necesidad de revisar la columna de x_j de la tabla símplex final (paso 2). De todas maneras se hará para ilustrarlo.

$$z_1^* - \bar{c}_1 = \mathbf{y}^*\bar{\mathbf{A}}_1 - c_1 = [0, 0, \tfrac{5}{2}]\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - 4 = 1.$$



■ **FIGURA 6.5**
Región factible de la variación 4 del modelo de la Wyndor, donde la variación 2 (figura 6.3) se ha modificado de modo que $a_{31} = 3 \rightarrow 2$ y $c_1 = 3 \rightarrow 4$.

$$\mathbf{A}_1^* = \mathbf{S}^*\overline{\mathbf{A}}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}.$$

El hecho de que $z_1^* - \bar{c}_1 \geq 0$ confirma de nuevo la optimalidad de la solución actual. Como $z_1^* - c_1$ es la variable de superávit de la restricción revisada del problema dual, esta prueba de optimalidad es equivalente a la que se usó antes.

Esto completa el análisis del efecto de los cambios del modelo actual (variación 2) a la variación 4. Debido a que otros cambios más grandes en las estimaciones originales de los coeficientes de x_1 serían poco realistas, el equipo de IO concluyó que estos coeficientes son parámetros *no sensibles* del modelo actual. Por tanto, se mantendrán fijos con el valor de sus mejores estimaciones que se muestran en la tabla 6.21 —$c_1 = 3$ y $a_{31} = 3$— para el resto del análisis de sensibilidad.

Intervalo permitido para un coeficiente de la función objetivo de una variable no básica. Se acaba de describir e ilustrar la forma de analizar cambios *simultáneos* en los coeficientes de una variable no básica x_j. Es una práctica común en el análisis de sensibilidad poner atención también en el efecto de cambiar *sólo un* parámetro, c_j. Esto incluye simplificar el enfoque anterior para encontrar el *intervalo de valores permitidos* para seguir óptima para c_j.

> Recuerde de la sección 4.7 que, para cualquier c_j, su intervalo permitido para permanecer óptima es el intervalo de valores para el que la solución óptima actual (que se obtuvo por el método símplex para el modelo actual antes de cambiar c_j) continúa siendo óptima. (Se supone que el cambio en esta única c_j es el único cambio del modelo actual.) Cuando x_j es una variable no básica para esta solución, la solución permanece óptima mientras $z_j^* - c_j \geq 0$, donde $z_j^* = \mathbf{y}^*\mathbf{A}_j$ es una constante a la que no afecta el cambio en el valor de c_j. En consecuencia, el intervalo permitido para c_j se puede calcular como $c_j \leq y^*\mathbf{A}_j$.

Por ejemplo, considere el modelo actual (variación 2) del problema de la Wyndor Glass Co., que se resume en el lado izquierdo de la tabla 6.21, donde la solución óptima actual (con $c_1 = 3$) está dada en el lado derecho. Al considerar sólo las variables de decisión, x_1 y x_2, la solución óptima es $(x_1, x_2) = (0, 9)$, como se ve en la figura 6.3. Cuando sólo se cambia c_1 esta solución permanece óptima siempre y cuando

$$c_1 \leq \mathbf{y}^*\mathbf{A}_1 = [0, 0, \tfrac{5}{2}] \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} = 7\tfrac{1}{2},$$

de manera que $c_1 \leq 7\frac{1}{2}$ es el intervalo permitido para permanecer óptima.

Una alternativa para realizar esta multiplicación de vectores es observar en la tabla 6.21 que $z_1^* - c_1 = \frac{9}{2}$ (el coeficiente de x_1 en el renglón 0) cuando $c_1 = 3$, de manera que $z_1^* = 3 + \frac{9}{2} = 7\frac{1}{2}$. Como $z_1^* = \mathbf{y}^*\mathbf{A}_1$, de inmediato se llega al mismo intervalo permisible.

La figura 6.3 proporciona un panorama gráfico de por qué $c_1 \leq 7\frac{1}{2}$ es el intervalo permisible. En $c_1 = 7\frac{1}{2}$, la función objetivo se convierte en $Z = 7.5x_1 + 5x_2 = 2.5(3x_1 + 2x_2)$, lo que quiere decir que la recta óptima del objetivo está arriba de la recta de frontera de restricción $3x_1 + 2x_2 = 18$ que se muestra en la figura. Así, en este punto terminal del intervalo permisible, se tiene una solución óptima múltiple que consiste en el segmento de recta entre (0, 9) y (4, 3). Si c_1 se aumentara más ($c_1 > 7\frac{1}{2}$), sólo (4, 3) sería óptimo. En consecuencia, se necesita que $c_1 \leq 7\frac{1}{2}$ para que (0, 9) siga óptimo.

El IOR Tutorial incluye un procedimiento llamado *Graphical Method and Sensitivity Analysis* (método gráfico y análisis de sensibilidad) que permite desarrollar este tipo de análisis gráfico de manera muy eficiente.

Para cualquier variable de decisión no básica x_j, en ocasiones se hace referencia al valor $z_j^* - c_j$ como el **costo reducido** de x_j, porque es la cantidad mínima en la que tendría que *reducirse* el *costo* unitario de la actividad *j* para hacer que valga la pena realizarla (aumentar el valor de x_j a más de cero). Cuando se interpreta c_j como la ganancia unitaria de la actividad *j* (lo que reduce el

incremento en el costo unitario c_j por la misma cantidad), el valor de $z_j^* - c_j$ es entonces el incremento máximo permitido de c_j para que la solución BF actual permancezca óptima.

La información que genera el análisis de sensibilidad de los paquetes de programación lineal, por lo general incluye tanto el costo reducido y el intervalo permisible para permanecer óptimo de cada coeficiente de la función objetivo (junto con el tipo de información que se despliega en la tabla 6.22). Esto se ilustró en las figuras 4.10 en el caso de Excel Solver y en las figuras A4.1 y A4.2 en los de LINGO y LINDO. La tabla 6.23 muestra esta información de la manera común para el modelo actual (variación 2 del modelo de la Wyndor). Las últimas tres columnas se usan para calcular el intervalo permisible para seguir óptima para cada coeficiente, los cuales son

$$c_1 \leq 3 + 4.5 = 7.5,$$
$$c_2 \geq 5 - 3 = 2.$$

Como se estudió en la sección 4.7, si cualquiera de los incrementos o decrementos hubiera sido cero, se hubiera tenido una indicación de que la solución óptima dada en la tabla es sólo una de las múltiples soluciones óptimas. En este caso, cambiar al coeficiente correspondiente una cantidad muy pequeña a más del cero permitido y resolver de nuevo.

Hasta aquí, se ha descrito cómo calcular el tipo de información de la tabla 6.23 sólo en el caso de variables no básicas. En el caso de una variable básica como x_2, el costo reducido automáticamente es 0. Se analizará cómo obtener el intervalo permisible para que c_j siga óptima cuando x_j es una variable básica en el caso 3.

Análisis de cambios simultáneos en los coeficientes de la función objetivo. Sin que importe si x_j es una variable básica o no básica, el intervalo permisible para seguir óptimo de c_j es válido sólo si este coeficiente de la función objetivo es el único que cambia. Sin embargo, cuando se hacen cambios simultáneos en los coeficientes de la función objetivo, se dispone de la regla de 100% para verificar si la solución original debe seguir óptima. De manera muy parecida a la regla de 100% de los cambios simultáneos en los lados derechos, esta regla combina los *cambios permisibles* (aumento o disminución) de las c_j individuales dadas en las dos últimas columnas de la tabla 6.23, como se describe en seguida.

> **Regla de 100% para cambios simultáneos en los coeficientes de la función objetivo:** Si se hacen cambios simultáneos en los coeficientes de la función objetivo, en cada cambio se calcula el porcentaje del cambio permisible (aumento o disminución) para que ese coeficiente siga dentro de su intervalo permisible para continuar óptimo. Si la *suma* de los porcentajes de cambios *no* excede 100%, la solución óptima original definitivamente será todavía óptima. (Si la suma excede de 100%, entonces no hay seguridad.)

Al usar la tabla 6.23 (y remitirse a la figura 6.3 para la visualización), esta regla de 100% indica que (0, 9) seguirá óptima para la variación 2 del modelo de la Wyndor Glass Co., aun cuando, al mismo tiempo, se aumente c_1 a más de 3 y se disminuya c_2 a menos de 5, siempre que estos cambios no sean muy grandes. Por ejemplo, si c_1 se incrementa 1.5 ($33\frac{1}{3}$% del cambio permisible), entonces c_2 puede disminuir 2 ($66\frac{2}{3}$% del cambio permisible). De manera similar, si c_1 aumenta 3 ($66\frac{2}{3}$% del cambio permisible), entonces c_2 puede disminuir sólo 1 ($33\frac{1}{3}$% del cambio permisible). Estos cambios máximos modifican la función objetivo ya sea como $Z = 4.5x_1 + 3x_2$ o bien $Z = 6x_1 + 4x_2$, con lo que la recta de la función objetivo óptima de la figura 6.3 rota en el sentido de las manecillas del reloj hasta que coincide con la ecuación de frontera de restricción $3x_1 + 2x_2 = 18$.

En general, cuando los coeficientes de la función objetivo cambian en la *misma* dirección, es posible que los porcentajes de los cambios permisibles sumen más de 100% sin cambiar la solución óptima. Se dará un ejemplo al final de la presentación del caso 3.

TABLA 6.23 Salida de software típica para el análisis de sensibilidad de los coeficientes de la función objetivo de la variación 2 del modelo de la Wyndor Glass Co.

Variable	Valor	Costo reducido	Coeficiente actual	Aumento permisible	Disminución permisible
x_1	0	4.5	3	4.5	∞
x_2	9	0	5	∞	3

Caso 2*b*: introducción de una nueva variable

Después de obtener la solución óptima se puede descubrir que la formulación de programación lineal no tomó en cuenta todas las actividades que pudieran ser atractivas. Para considerar una nueva actividad se requiere introducir una nueva variable con los coeficientes apropiados en la función objetivo y en las restricciones del modelo actual, que es la situación que se presenta en el caso 2*b*.

La forma conveniente de manejar este caso es tratarlo como si fuera el caso 2*a*. Para realizar este procedimiento se presume que la nueva variable x_j en realidad formaba parte del modelo original con todos sus coeficientes iguales a cero (por lo que todavía son cero en la tabla símplex final) y que x_j es una variable no básica de la solución BF actual. Entonces, si se cambian estos coeficientes de cero a sus valores reales para la nueva variable, sin duda el procedimiento (que incluye la reoptimización) se vuelve idéntico al del caso 2*a*.

En particular, todo lo que se debe hacer para comprobar si la solución actual es todavía óptima es verificar si la solución básica complementaria $\mathbf{y}^*$ satisface la nueva restricción dual que corresponde a la nueva variable del problema primal. Este enfoque se describió y después se ejemplificó para el problema de la Wyndor Glass Co., en la sección 6.5.

Caso 3: cambios en los coeficientes de una variable básica

Ahora suponga que la variable x_j (con j fija) que está en estudio es una variable *básica* de la solución óptima que se muestra en la tabla símplex final. En el caso 3 se supone que los únicos cambios al modelo actual se hacen en los coeficientes de esta variable.

El caso 3 difiere del 2*a* debido al requisito de que la tabla símplex debe estar en la forma apropiada de eliminación de Gauss. Esta forma permite que los elementos de la columna de una variable no básica tengan cualquier valor, así que no afecta en el caso 2*a*. Sin embargo, en el caso 3 la variable básica x_j debe tener coeficiente 1 en su renglón de la tabla símplex y coeficiente 0 en todos los demás renglones (incluso en el renglón 0). Por tanto, una vez calculados los cambios en la columna x_j de la tabla símplex final,[6] es probable que sea necesario aplicar la eliminación de Gauss para restaurar la forma apropiada, como se ejemplificó en la tabla 6.20. Este paso, a su vez, quizá cambie los valores de la solución básica actual, y puede hacerla no factible o no óptima (con lo que tal vez sea necesario reoptimizar). En consecuencia, el caso 3 requiere todos los pasos del procedimiento general que se resume al final de la sección 6.6.

Antes de aplicar la eliminación de Gauss, las fórmulas para verificar la columna de x_j son las mismas que para el caso 2*a*, y se resumen de la siguiente forma.

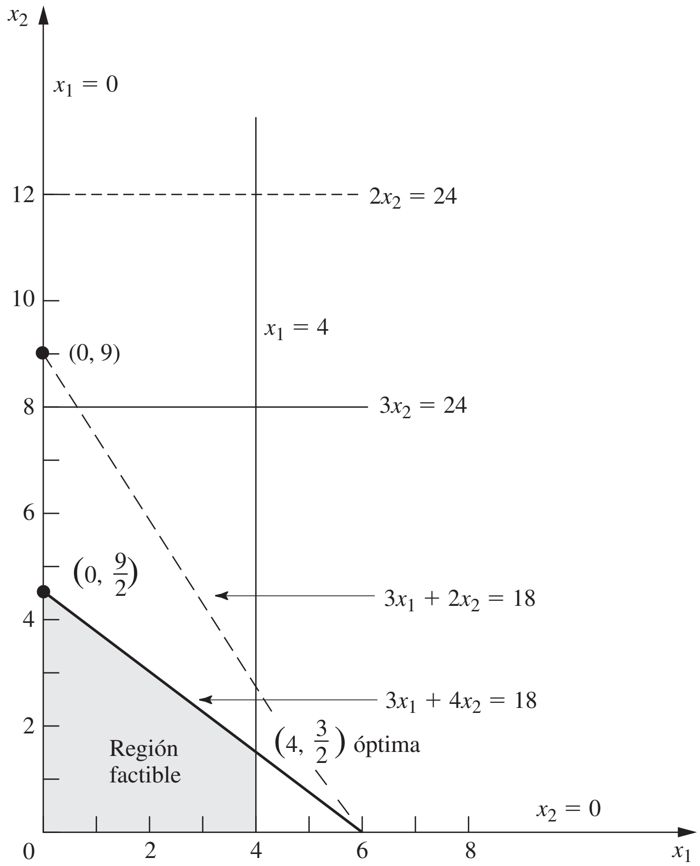
Coeficiente de x_j del renglón 0 final: $z_j^* - \bar{c}_j = \mathbf{y}^*\bar{\mathbf{A}}_j - \bar{c}_j$.

Coeficiente de x_j de los renglones 1 a m finales: $\mathbf{A}_j^* = \mathbf{S}^*\bar{\mathbf{A}}_j$.

Ejemplo (variación 5 del modelo de la Wyndor). Como x_2 es una variable básica de la tabla 6.21 para la variación 2 del modelo de la Wyndor Glass Co., el análisis de sensibilidad de sus coeficientes se ajusta al caso 3. Dada la solución óptima actual ($x_1 = 0, x_2 = 9$), el producto 2 es el único producto nuevo que debe introducirse, y su tasa de producción será relativamente grande. Por ello, la pregunta clave es si las estimaciones iniciales que condujeron a los coeficientes de x_2 en el modelo actual (variación 2) pudieron haber *sobrestimado* tanto las cualidades del producto 2 como para que invaliden esta conclusión. Para responder a esta pregunta se debe verificar el conjunto *más pesimista* de estimaciones razonables para estos coeficientes, que resulta ser $c_2 = 3$, $a_{22} = 3$ y $a_{32} = 4$. En consecuencia, los cambios que deben investigarse (variación 5 del modelo Wyndor) son

$$c_2 = 5 \longrightarrow \bar{c}_2 = 3, \qquad \mathbf{A}_2 = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix} \longrightarrow \bar{\mathbf{A}}_2 = \begin{bmatrix} 0 \\ 3 \\ 4 \end{bmatrix}.$$

[6] Para el lector sutil, debe hacerse notar que en este punto, en el caso 3 existe la posibilidad de encontrarse con un obstáculo. Los cambios en la tabla inicial pueden destruir la independencia lineal de las columnas de los coeficientes de las variables básicas. Este hecho ocurre sólo si el coeficiente unitario de la variable básica x_j en la tabla final se cambia a cero, en cuyo caso será necesario realizar cálculos más extensos del método símplex para el caso 3.



■ **FIGURA 6.6**
Región factible de la variación 5 del modelo de la Wyndor, donde la variación 2 (figura 6.3) se ha modificado de modo que $c_2 = 5 \to 3$, $a_{22} = 2 \to 3$ y $a_{32} = 2 \to 4$.

El efecto gráfico de estos cambios es la modificación de la región factible de acuerdo con la figura 6.3 respecto de la que se muestra en la figura 6.6. La solución óptima en la figura 6.3 es $(x_1, x_2) = (0, 9)$, que corresponde a la solución en el vértice en donde se cruzan las fronteras de restricción $x_1 = 0$ y $3x_1 + 2x_2 = 18$. Cuando se modifican las restricciones, la solución en un vértice correspondiente en la figura 6.6 es $(0, \frac{9}{2})$. No obstante, esta solución ya no es óptima, puesto que la función objetivo revisada, $Z = 3x_1 + 3x_2$ conduce ahora a la nueva solución óptima de $(x_1, x_2) = (4, \frac{3}{2})$.

Análisis de la variación 5. Ahora se verá cómo se puede llegar a estas mismas conclusiones de manera algebraica. Dado que los únicos cambios en el modelo ocurren en los coeficientes de x_2, las *únicas* modificaciones que resultan en la tabla símplex final (tabla 6.21) están en la columna de x_2. En consecuencia, se usan las fórmulas anteriores para recalcular nada más esta columna.

$$z_2 - \bar{c}_2 = \mathbf{y}^*\bar{\mathbf{A}}_2 - \bar{c}_2 = [0, 0, \tfrac{5}{2}]\begin{bmatrix}0\\3\\4\end{bmatrix} - 3 = 7.$$

$$\mathbf{A}_2^* = \mathbf{S}^*\bar{\mathbf{A}}_2 = \begin{bmatrix}1 & 0 & 0\\0 & 0 & \frac{1}{2}\\0 & 1 & -1\end{bmatrix}\begin{bmatrix}0\\3\\4\end{bmatrix} = \begin{bmatrix}0\\2\\-1\end{bmatrix}.$$

(De manera equivalente, se puede usar el análisis incremental con $\Delta c_2 = -2$, $\Delta a_{22} = 1$ y $\Delta a_{32} = 2$ de la misma forma para obtener esta columna.)

La tabla final revisada que resulta se muestra en la parte superior de la tabla 6.24. Observe que los nuevos coeficientes de esta variable básica x_2 no tienen los valores requeridos y se tiene que aplicar la conversión a la forma apropiada con eliminación de Gauss. Este paso exige dividir

TABLA 6.24 Procedimiento del análisis de sensibilidad aplicado a la variación 5 del modelo de la Wyndor Glass Co.

	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	x_5	Lado derecho
Tabla revisada final	Z	(0)	1	$\frac{9}{2}$	7	0	0	$\frac{5}{2}$	45
	x_3	(1)	0	1	0	1	0	0	4
	x_2	(2)	0	$\frac{3}{2}$	2	0	0	$\frac{1}{2}$	9
	x_4	(3)	0	−3	−1	0	1	−1	6
Convertida a la forma apropiada	Z	(0)	1	$-\frac{3}{4}$	0	0	0	$\frac{3}{4}$	$\frac{27}{2}$
	x_3	(1)	0	1	0	1	0	0	4
	x_2	(2)	0	$\frac{3}{4}$	1	0	0	$\frac{1}{4}$	$\frac{9}{2}$
	x_4	(3)	0	$-\frac{9}{4}$	0	0	1	$-\frac{3}{4}$	$\frac{21}{2}$
Nueva tabla final después de la reoptimización (en este caso sólo se requirió una iteración del método símplex)	Z	(0)	1	0	0	$\frac{3}{4}$	0	$\frac{3}{4}$	$\frac{33}{2}$
	x_1	(1)	0	1	0	1	0	0	4
	x_2	(2)	0	0	1	$-\frac{3}{4}$	0	$\frac{1}{4}$	$\frac{3}{2}$
	x_4	(3)	0	0	0	$\frac{9}{4}$	1	$-\frac{3}{4}$	$\frac{39}{2}$

el renglón 2 entre 2, restar el nuevo renglón 2 multiplicado por 7 del renglón 0 y sumar el nuevo renglón 2 al renglón 3.

La segunda tabla símplex de la tabla 6.24 da los nuevos valores de la solución básica actual, es decir, $x_3 = 4$, $x_2 = \frac{9}{2}$, $x_4 = \frac{21}{2}$ ($x_1 = 0$, $x_5 = 0$). Como todas estas variables son no negativas, la solución todavía es factible. Sin embargo, el coeficiente negativo de x_1 en el renglón 0 indica que la solución ya no es óptima. Se aplicará el método símplex a esta tabla, con esta solución como solución BF inicial, para encontrar la nueva solución óptima. La variable entrante básica inicial es x_1, con x_3 como la variable básica que sale. En este caso se necesita sólo una iteración para llegar a la nueva solución óptima: $x_1 = 4$, $x_2 = \frac{3}{2}$, $x_4 = \frac{39}{2}$ ($x_3 = 0$, $x_5 = 0$), como se muestra en la tabla 6.24.

Todo este análisis sugiere que c_2, a_{22} y a_{32} son parámetros relativamente sensibles. Sin embargo, para estimar con más cuidado, los datos adicionales sólo pueden obtenerse si se realiza una prueba piloto. Por tanto, el equipo de IO recomienda que se inicie de inmediato la producción del artículo 2 en pequeña escala ($x_2 = \frac{3}{2}$) y que se use esta experiencia como guía para decidir si la capacidad de producción restante debe asignarse al producto 2 o al producto 1.

Intervalo permitido del coeficiente de la función objetivo de una variable básica. En el caso 2*a* se describió cómo encontrar el intervalo permisible para que cualquier c_j siguiera óptima tal que x_j sea una variable no básica para la solución óptima actual (antes de cambiar c_j). Cuando x_j es una variable básica, el procedimiento se complica un poco por la necesidad de convertir a la forma apropiada de eliminación de Gauss antes de probar la optimalidad.

Para ilustrar el procedimiento, considere la variación 5 del modelo de la Wyndor Glass Co. (con $c_2 = 3$, $a_{22} = 3$, $a_{23} = 4$) cuya gráfica se presenta en la figura 6.6 y que se resuelve en la tabla 6.24. Como x_2 es una variable básica de la solución óptima dada al final de la tabla (con $c_2 = 3$) los pasos necesarios para encontrar el intervalo de valores permitidos para que c_2 siga óptima son los siguientes:

1. Dado que x_2 es una variable básica, observe que su coeficiente en el nuevo renglón 0 final (vea la parte inferior de la tabla símplex en la tabla 6.24) de modo automático es $z_2^* - c_2 = 0$ antes de cambiar c_2 de su valor actual de 3.

2. Ahora se incrementa $c_2 = 3$ en Δc_2 (de manera que $c_2 = 3 + \Delta c_2$). Esto cambia el coeficiente indicado en el paso 1 a $z_2^* - c_2 = -\Delta c_2$, lo que cambia el renglón 0 a

$$\text{Renglón } 0 = \left[0,\ -\Delta c_2,\ \frac{3}{4},\ 0,\ \frac{3}{4}\ \vdots\ \frac{33}{2}\right].$$

3. Con este coeficiente que ahora es diferente de cero, deben realizarse operaciones elementales para restaurar la forma apropiada de eliminación de Gauss. En particular, se suma al renglón 0 el renglón 2 multiplicado por Δc_2, de donde se obtiene el nuevo renglón cero, como aparece a continuación.

$$\begin{array}{rl} & \left[0,\ -\Delta c_2,\ \frac{3}{4},\ 0,\ \frac{3}{4}\ \vdots\ \frac{33}{2}\right] \\ + & \left[0,\ \Delta c_2,\ -\frac{3}{4}\Delta c_2,\ 0,\ \frac{1}{4}\Delta c_2\ \vdots\ \frac{3}{2}\Delta c_2\right] \\ \hline \text{Nuevo renglón } 0 = & \left[0,\ 0,\ \frac{3}{4} - \frac{3}{4}\Delta c_2,\ 0,\ \frac{3}{4} + \frac{1}{4}\Delta c_2\ \vdots\ \frac{33}{2} + \frac{3}{2}\Delta c_2\right] \end{array}$$

4. Si se usa este nuevo renglón 0, se calcula el intervalo de valores de Δc_2 que mantiene no negativos a los coeficientes de las variables no básicas (x_3 y x_5).

$$\frac{3}{4} - \frac{3}{4}\Delta c_2 \geq 0 \quad \Rightarrow \quad \frac{3}{4} \geq \frac{3}{4}\Delta c_2 \quad \Rightarrow \quad \Delta c_2 \leq 1.$$

$$\frac{3}{4} + \frac{1}{4}\Delta c_2 \geq 0 \quad \Rightarrow \quad \frac{1}{4}\Delta c_2 \geq -\frac{3}{4} \quad \Rightarrow \quad \Delta c_2 \geq -3.$$

Entonces, el intervalo de valores es $-3 \leq \Delta c_2 \leq 1$.

5. Como $c_2 = 3 + \Delta c_2$, se suma 3 a este intervalo de valores, de donde se obtiene

$$0 \leq c_2 \leq 4$$

como el intervalo de valores permitido para que c_2 permanezca óptima.

Con sólo dos variables de decisión, este intervalo permitido se puede verificar en una gráfica mediante la figura 6.6 con una función objetivo $Z = 3x_1 + c_2x_2$. Con el valor actual $c_2 = 3$, la solución óptima es $(4, \frac{3}{2})$. Cuando se incrementa c_2 esta solución sigue óptima sólo para $c_2 \leq 4$. Para $c_2 \geq 4$, $(4, \frac{9}{2})$ se convierte en óptima (con un empate en $c_2 = 4$), debido a la frontera de restricción $3x_1 + 4x_2 = 18$. Cuando, por el contrario, c_2 disminuye, $(4, \frac{3}{2})$ sigue óptima sólo para $c_2 \geq 0$. Si $c_2 \leq 0$, $(4, 0)$ se vuelve óptima debido a la frontera de restricción $x_1 = 4$.

De manera similar, el intervalo de valores permitidos para permanecer óptima para c_1 (con c_2 fijo en 3) se puede obtener algebraica o gráficamente como $c_1 \geq \frac{9}{4}$. (El problema 6.7-10 le pide que verifique este resultado de ambas formas.)

De esta forma, la *disminución permisible* de c_1 a partir del valor actual de 3 es de sólo $\frac{3}{4}$. Sin embargo, es posible disminuir c_1 una cantidad mayor sin cambiar la solución óptima si c_2 también disminuye lo suficiente. Por ejemplo, suponga que *ambos*, c_1 y c_2, se disminuyen 1 unidad a partir de su valor actual de 3, de manera que la función objetivo cambia de $Z = 3x_1 + 3x_2$ a $Z = 2x_1 + 2x_2$. De acuerdo con la regla de 100% de los cambios simultáneos en los coeficientes de la función objetivo, los porcentajes de cambios permisibles son $133\frac{1}{3}$% y $33\frac{1}{3}$%, respectivamente, que suman mucho más que 100%. No obstante, la pendiente de la recta de la función objetivo no cambió en absoluto, por lo cual $(4, \frac{3}{2})$ todavía es óptima.

Caso 4: introducción de una nueva restricción

El último caso es aquel en el que debe introducirse al modelo una nueva restricción, después de obtener la solución. Este caso puede ocurrir porque se pasó por alto la restricción en un principio o porque surgieron nuevas consideraciones después de formular el modelo. Otra posibilidad es que se haya eliminado, a propósito, la restricción para disminuir el esfuerzo computacional porque parece

ser menos restrictiva que otras ya planteadas en el modelo, pero ahora es necesario verificar esta impresión con la solución óptima obtenida.

Para ver si la nueva restricción afecta a la solución óptima actual, todo lo que debe hacerse es verificar en forma directa si esa solución óptima satisface la restricción. Si es así, todavía sería la *mejor solución básica factible* (es decir, la solución óptima), aun cuando se agregara la restricción al modelo. La razón es que una nueva restricción sólo puede eliminar algunas de las soluciones factibles anteriores sin agregar algunas nuevas.

Si la nueva restricción elimina la solución óptima actual, y si se quiere encontrar la nueva solución, se introduce esta restricción a la tabla símplex final (como un renglón adicional) exactamente como si fuera la tabla inicial, en la que se designa la variable usual (de holgura o artificial) como la variable básica que corresponde a este nuevo renglón. Como éste tal vez tenga coeficientes *distintos de cero* en algunas otras variables básicas, se debe aplicar la conversión a la forma apropiada de eliminación de Gauss y después el paso de reoptimización en la forma normal.

Igual que para algunos de los casos anteriores, el procedimiento del caso 4 es una versión simplificada del procedimiento general que se resume al final de la sección 6.6. La única pregunta que hay que hacerse en este caso es si la solución óptima anterior todavía es *factible*, así que debe eliminarse el paso 5 (prueba de optimalidad). El paso 4 (prueba de factibilidad) se reemplaza por una prueba de factibilidad mucho más rápida (¿La solución óptima anterior satisface la nueva restricción?) que se realiza inmediatamente después del paso 1 (revisión del modelo). Sólo cuando la respuesta a esta prueba es negativa y se quiere reoptimizar, se cubren los pasos 2, 3 y 6 (revisión de la tabla símplex final, conversión a la forma apropiada de eliminación de Gauss y reoptimización).

Ejemplo (variación 6 del modelo de la Wyndor). Para ilustrar este caso, se considera la variación 6 del modelo de la Wyndor Glass Co., que introduce la nueva restricción

$$2x_1 + 3x_2 \leq 24$$

en la variación 2 del modelo dado en la tabla 6.21. El efecto gráfico se muestra en la figura 6.7. La solución óptima anterior (0, 9) viola la nueva restricción, por lo que la solución óptima cambia a (0, 8).

Para analizar este ejemplo en forma algebraica, observe que con (0, 9) se obtiene $2x_1 + 3x_2 = 27 > 24$, por lo que esta solución óptima anterior ya no es factible. Para encontrar la nueva solución óptima, se agrega esta restricción a la tabla símplex final actual como se describió, con la variable de holgura x_6 como su variable básica inicial. Este paso conduce a la primera tabla que se muestra en la tabla 6.25. El paso de conversión a la forma apropiada de eliminación de Gauss requiere restar el renglón 2 multiplicado por 3, del nuevo renglón, con lo que se identifica la solución básica actual: $x_3 = 4, x_2 = 9, x_4 = 6, x_6 = -3$ ($x_1 = 0, x_5 = 0$), como se muestra en la segunda tabla símplex. Cuando se aplica el método dual símplex (descrito en la sección 7.1) se obtiene en una sola iteración (algunas veces se necesitan más) la nueva solución óptima en la tabla final de la tabla 6.25.

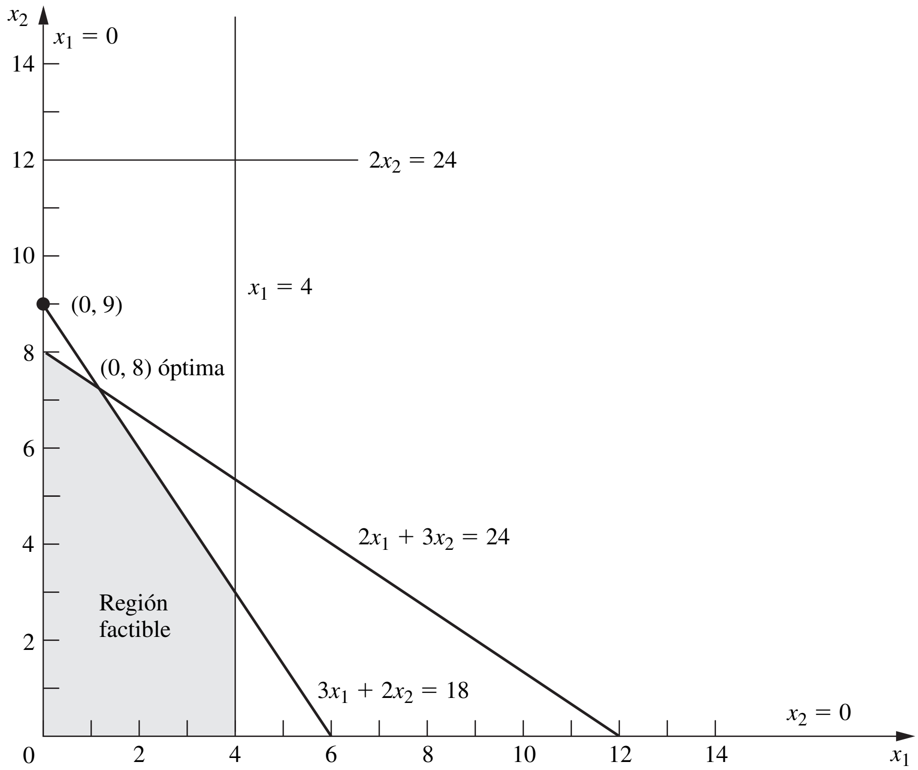
Análisis de sensibilidad sistemático: programación paramétrica

Hasta aquí se ha descrito cómo verificar cambios específicos en los parámetros del modelo. Otro enfoque común del análisis de sensibilidad es modificar de manera continua uno o más parámetros en un intervalo (o intervalos) para ver cuándo cambia la solución óptima.

Por ejemplo, en la variación 2 del modelo de la Wyndor Glass Co., en lugar de comenzar a probar el cambio específico de $b_2 = 12$ a $\bar{b}_2 = 24$, se puede establecer

$$\bar{b}_2 = 12 + \theta$$

y después modificar θ de manera continua desde 0 hasta 12 (valor máximo de interés). La interpretación geométrica de la figura 6.3 es que la recta de restricción $2x_2 = 12$ se mueve hacia arriba hasta $2x_2 = 12 + \theta$, donde el valor de θ aumenta de 0 a 12. El resultado es que la solución FEV óptima original (2, 6) mueve la recta de restricción $3x_1 + 2x_2 = 18$ hacia el punto $(-2, 12)$. Esta solución en un vértice permanece óptima mientras siga siendo factible ($x_1 \geq 0$), después de lo cual (0, 9) se convierte en la solución óptima.



■ **FIGURA 6.7**
Región factible de la variación 6 del modelo de la Wyndor, donde la variación 2 (figura 6.3) se ha modificado por la adición de una nueva restricción $2x_1 + 3x_2 \leq 24$.

Los cálculos algebraicos del efecto de tener $\Delta b_2 = \theta$ son directamente análogos a los del ejemplo del caso 1 en donde $\Delta b_2 = 12$. En particular, si se emplean las expresiones de Z^* y $\mathbf{b}^*$ dadas para el caso 1,

$$Z^* = \mathbf{y}^*\overline{\mathbf{b}}$$
$$\mathbf{b}^* = \mathbf{S}^*\overline{\mathbf{b}}$$

donde ahora es

$$\overline{\mathbf{b}} = \begin{bmatrix} 4 \\ 12+\theta \\ 18 \end{bmatrix}$$

y donde $\mathbf{y}^*$ y $\mathbf{S}^*$ se presentan en los cuadros de la tabla símplex intermedia de la tabla 6.19. Estas ecuaciones indican que la solución óptima es

$$Z^* = 36 + \frac{3}{2}\theta$$
$$x_3 = 2 + \frac{1}{3}\theta \qquad\qquad (x_4 = 0,\ x_5 = 0)$$
$$x_2 = 6 + \frac{1}{2}\theta$$
$$x_1 = 2 - \frac{1}{3}\theta$$

para valores de θ tan pequeños como para que esta solución siga factible, esto es, para $\theta \leq 6$. Para valores de $\theta > 6$, el método dual símplex (descrito en la sección 7.1) conduce a la tabla que se muestra en la tabla 6.21 excepto por el valor de x_4. Así $Z = 45$, $x_3 = 4$, $x_2 = 9$ (junto con $x_1 = 0$, $x_5 = 0$), y de la expresión de $\mathbf{b}^*$ se obtiene

$$x_4 = b_3^* = 0(4) + 1(12 + \theta) - 1(18) = -6 + \theta.$$

TABLA 6.25 Procedimiento del análisis de sensibilidad aplicado a la variación 6 del modelo de la Wyndor Glass Co.

	Variable básica	Ec.	Z	x_1	x_2	x_3	x_4	x_5	x_6	Lado derecho
Tabla revisada final	Z	(0)	1	$\frac{9}{2}$	0	0	0	$\frac{5}{2}$	0	45
	x_3	(1)	0	1	0	1	0	0	0	4
	x_2	(2)	0	$\frac{3}{2}$	1	0	0	$\frac{1}{2}$	0	9
	x_4	(3)	0	-3	0	0	1	-1	0	6
	x_6	Nueva	0	2	3	0	0	0	1	24
Convertida a la forma apropiada	Z	(0)	1	$\frac{9}{2}$	0	0	0	$\frac{5}{2}$	0	45
	x_3	(1)	0	1	0	1	0	0	0	4
	x_2	(2)	0	$\frac{3}{2}$	1	0	0	$\frac{1}{2}$	0	9
	x_4	(3)	0	-3	0	0	1	-1	0	6
	x_6	Nueva	0	$-\frac{5}{2}$	0	0	0	$-\frac{3}{2}$	1	-3
Nueva tabla final después de la reoptimización (en este caso sólo se requirió una iteración del método símplex)	Z	(0)	1	$\frac{1}{3}$	0	0	0	0	$\frac{5}{3}$	40
	x_3	(1)	0	1	0	1	0	0	0	4
	x_2	(2)	0	$\frac{2}{3}$	1	0	0	0	$\frac{1}{3}$	8
	x_4	(3)	0	$-\frac{4}{3}$	0	0	1	0	$-\frac{2}{3}$	8
	x_5	Nueva	0	$\frac{5}{3}$	0	0	0	1	$-\frac{2}{3}$	2

Después, esta información se puede usar (junto con otros datos no incorporados al modelo sobre el efecto de incrementar b_2) para decidir si se debe conservar la solución óptima original o, en caso contrario, cuánto debe aumentar el valor de b_2.

De manera similar se puede investigar el efecto que tiene, en la solución óptima, modificar más de un parámetro a la vez. Cuando se modifican sólo los parámetros b_i, se expresa el nuevo valor b_i en términos del valor original b_i de la siguiente manera:

$$\bar{b}_i = b_i + \alpha_i\theta, \qquad \text{para } i = 1, 2, \ldots, m,$$

donde los valores de α_i son constantes de entrada que especifican la tasa de incremento deseada (positiva o negativa) de los lados derechos correspondientes, al aumentar el valor de θ.

Por ejemplo, suponga que es posible trasladar cierta parte de la fabricación de un producto actual de la Wyndor Glass Co., de la planta 2 a la 3, si se aumenta b_3 y se disminuye b_2. Además, suponga que b_3, disminuye el doble de lo que crece b_2. Entonces,

$$\bar{b}_2 = 12 + \theta$$
$$\bar{b}_3 = 18 - 2\theta,$$

donde el valor de θ (no negativo) mide la cantidad de producción transferida. (Así, en este caso, $\alpha_1 = 0$, $\alpha_2 = 1$ y $\alpha_3 = -2$.) En la figura 6.3, la interpretación geométrica es que conforme θ aumenta su valor, la recta de restricción $2x_2 = 12$ se mueve hacia arriba hasta $2x_2 = 12 + \theta$ (ignore la recta $2x_2 = 24$) y en forma simultánea la recta de restricción $3x_1 + 2x_2 = 18$ se mueve hacia abajo hasta $3x_1 + 2x_2 = 18 - 2\theta$. La solución FEV óptima original (2, 6) se encuentra en el cruce de las rectas $2x_2 = 12$ y $3x_1 + 2x_2 = 18$, así que al modificarse estas rectas, la solución en un vértice cambia. No obstante, con la función objetivo $Z = 3x_1 + 5x_2$, la solución en el vértice permanecerá óptima mientras sea factible ($x_1 \geq 0$).

Una investigación algebraica de estos cambios simultáneos en b_2 y b_3 implica de nuevo usar las fórmulas para el caso 1 (donde θ representa un número desconocido) para calcular los cambios resultantes en la tabla símplex final (parte media de la tabla 6.19), es decir,

$$Z^* = \mathbf{y}^*\overline{\mathbf{b}} = [0, \tfrac{3}{2}, 1]\begin{bmatrix} 4 \\ 12 + \theta \\ 18 - 2\theta \end{bmatrix} = 36 - \tfrac{1}{2}\theta,$$

$$\mathbf{b}^* = \mathbf{S}^*\overline{\mathbf{b}} = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 4 \\ 12 + \theta \\ 18 - 2\theta \end{bmatrix} = \begin{bmatrix} 2 + \theta \\ 6 + \frac{1}{2}\theta \\ 2 - \theta \end{bmatrix}.$$

Por tanto, la solución óptima se convierte en

$$Z^* = 36 - \frac{1}{2}\theta$$

$$x_3 = 2 + \theta \qquad (x_4 = 0, \qquad x_5 = 0)$$

$$x_2 = 6 + \frac{1}{2}\theta$$

$$x_1 = 2 - \theta$$

para θ tan pequeño como para que esta solución siga factible, por ejemplo, para $\theta \leq 2$. (Verifique esta conclusión en la figura 6.3.) Sin embargo, el hecho de que Z disminuya su valor al aumentar el de θ desde 0 indica que la mejor elección para θ es $\theta = 0$, así que no debe hacerse ningún cambio en la producción.

El enfoque para cambiar varios parámetros c_j al mismo tiempo es similar. En este caso, se expresa el nuevo valor de c_j en términos del valor original de c_j de la siguiente manera

$$\overline{c}_j = c_j + \alpha_j\theta, \qquad \text{para } j = 1, 2, \ldots, n,$$

donde las α_j son constantes de entrada que especifican la tasa deseada de incremento (positiva o negativa) de c_j al aumentar el valor de θ.

Para ilustrar este caso, reconsidere el análisis de sensibilidad de c_1 y c_2 del problema de la Wyndor Glass Co., que se realizó antes en esta sección. Se comenzó con la variación 2 del modelo de la Wyndor que se presenta en la tabla 6.21 y la figura 6.3, y se estudió por separado el efecto de cambiar c_1 de 3 a 4 (su estimación más optimista) y c_2 de 5 a 3 (su estimación más pesimista). Ahora se pueden considerar ambos cambios al mismo tiempo, al igual que varios casos intermedios con cambios más pequeños, si se establece

$$\overline{c}_1 = 3 + \theta \qquad \text{y} \qquad \overline{c}_2 = 5 - 2\theta,$$

donde el valor de θ mide la *fracción* del máximo cambio posible que se hace. El resultado es reemplazar la función objetivo original, $Z = 3x_1 + 5x_2$ por una *función* de θ

$$Z(\theta) = (3 + \theta)x_1 + (5 - 2\theta)x_2,$$

de forma que ahora se puede realizar la optimización para cualquier valor deseado (fijo) de θ entre 0 y 1. Cuando se verifica el efecto del aumento de θ de 0 a 1, se puede determinar con exactitud cuándo y cómo cambia la solución óptima cuando aumenta el error en las estimaciones de estos parámetros originales.

La consideración de estos cambios simultáneos es especialmente apropiada si existen factores que provocan que los parámetros cambien al mismo tiempo. ¿Son competitivos los dos productos en algún sentido, de forma que una ganancia unitaria más grande que la esperada de uno de ellos implique una ganancia por unidad menor que la esperada para el otro? ¿Están los dos productos afectados por algún factor externo como una campaña publicitaria de un competidor? ¿Es posible cambiar de manera simultánea ambas ganancias unitarias mediante cambios apropiados de personal y equipo?

En la región factible de la figura 6.3, la interpretación geométrica de cambiar la función objetivo de $Z = 3x_1 + 5x_2$ a $Z(\theta) = (3 + \theta)x_1 + (5 - 2\theta)x_2$ es que cambia la *pendiente* de la recta de la

función objetivo original ($Z = 45 = 3x_1 + 5x_2$) que pasa por la solución óptima (0, 9). Si el valor de θ aumenta lo suficiente, esta pendiente cambiará hasta cambiar la solución óptima de (0, 9) a otra solución FEV (4, 3). (Verifique esto de manera gráfica para $\theta \leq 1$.)

El procedimiento algebraico para manejar estos dos cambios simultáneos ($\Delta c_1 = \theta$ y $\Delta c_2 = -2\theta$) se muestra en la tabla 6.26. Aunque los cambios ahora se expresan en términos de θ en lugar de ser cantidades numéricas específicas, θ se maneja igual que un número desconocido. La tabla muestra sólo los renglones relevantes de la tabla símplex (el renglón 0 y el renglón de la variable básica x_2). La primera tabla símplex que se mostró es exactamente la tabla símplex final de la versión actual del modelo (antes de cambiar c_1 y c_2) de acuerdo con la tabla 6.21. Según las fórmulas de la tabla 6.17, el único cambio en la tabla símplex final *revisada* que se muestra en seguida es Δc_1 y Δc_2 que se restan de los coeficientes de x_1 y x_2, del renglón 0, respectivamente. Para convertir esta tabla símplex en la forma apropiada de eliminación de Gauss se resta del renglón 0 el renglón 2 multiplicado por 2θ, lo que conduce a la última tabla símplex ilustrada. Las expresiones en términos de θ de los coeficientes de las variables no básicas x_1 y x_5 en el renglón 0 de esta tabla indican que la solución BF actual aún es óptima para $\theta < \frac{9}{8}$. Como $\theta = 1$ es el máximo valor realista de θ, se puede decir que c_1 y c_2 juntos son parámetros no sensibles respecto de la variación 2 del modelo en la tabla 6.21. No hay necesidad de tratar de estimar el valor de estos parámetros a menos que otros parámetros cambien (como ocurre con la variación 5 del modelo de la Wyndor).

Como ya se explicó en la sección 4.7, esta manera de modificar en forma continua algunos parámetros al mismo tiempo se conoce con el nombre de *programación lineal paramétrica*. En la sección 7.2 se presenta el procedimiento completo (que incluye la identificación de nuevas soluciones óptimas para valores más grandes de θ) cuando sólo varían los parámetros c_j y después cuando sólo se modifican los parámetros b_i. Algunos paquetes de computadora para programación lineal incluyen rutinas para modificar los coeficientes de una sola variable o los parámetros de una sola restricción. Además de las aplicaciones que se mencionaron en la sección 4.7, este procedimiento proporciona una forma conveniente de realizar un análisis de sensibilidad sistemático.

6.8 REALIZACIÓN DE ANÁLISIS DE SENSIBILIDAD EN UNA HOJA DE CÁLCULO[7]

Con la ayuda del Excel Solver, las hojas de cálculo proporcionan una alternativa, relativamente directa, para realizar gran parte del análisis de sensibilidad descrito en la secciones 6.5 a 6.7. El

TABLA 6.26 **Manejo de $\Delta c_1 = \theta$ y $\Delta c_2 = -2\theta$ de la variación 2 del modelo de la Wyndor dado en la tabla 6.21**

	Variable básica	Ec.	Z	Coeficiente de: x_1	x_2	x_3	x_4	x_5	Lado derecho
Tabla final	Z	(0)	1	$\frac{9}{2}$	0	0	0	$\frac{5}{2}$	45
	x_2	(2)	0	$\frac{3}{2}$	1	0	0	$\frac{1}{2}$	9
Tabla revisada final cuando $\Delta c_1 = \theta$ y $\Delta c_2 = -2\theta$	$Z(\theta)$	(0)	1	$\frac{9}{2} - \theta$	2θ	0	0	$\frac{5}{2}$	45
	x_2	(2)	0	$\frac{3}{2}$	1	0	0	$\frac{1}{2}$	9
Convertida a la forma apropiada	$Z(\theta)$	(0)	1	$\frac{9}{2} - 4\theta$	0	0	0	$\frac{5}{2} - \theta$	$45 - 18\theta$
	x_2	(2)	0	$\frac{3}{2}$	1	0	0	$\frac{1}{2}$	9

[7] Esta sección se escribió de forma que pueda ser entendida sin antes haber leído las secciones anteriores a este capítulo. Sin embargo, la sección 4.7 es una base importante para la última parte de esta sección.

enfoque de la hoja de cálculo es en esencia el mismo para cada uno de los casos que se consideran en la sección 6.7 de los tipos de cambios que se introdujeron al modelo original. Por tanto, se dará máxima relevancia al efecto de los cambios en los coeficientes de las variables de la función objetivo (casos 2*a* y 3 en la sección 6.7). Se ilustrará este efecto al hacer cambios en el modelo *original* de la Wyndor que se formuló en la sección 3.1, donde los coeficientes de x_1 (número de lotes de la puerta nueva producidos por semana) y x_2 (número de lotes de la ventana nueva producidos por semana) de la función objetivo son

$c_1 = 3 =$ ganancia (en miles de dólares) por lote del nuevo tipo de puerta,
$c_2 = 5 =$ ganancia (en miles de dólares) por lote del nuevo tipo de ventana.

Por conveniencia, la formulación en hoja de cálculo de este modelo (figura 3.22) se repite aquí como la figura 6.8. Observe que las celdas que contienen las cantidades que cambian son GananciaPorLote (C4:D4). Como las ganancias en estas celdas se expresan en dólares, mientras que c_1 y c_2 están en unidades de miles de dólares, de aquí en adelante se explicará el análisis de sensibilidad en términos de los cambios que se realizaron en estas celdas en vez de hablar de los cambios en c_1 y c_2. Con este propósito, estas ganancias se denotarán por

$G_P =$ Ganancia por lote de puertas que actualmente está en la celda C4,
$G_V =$ Ganancia por lote de ventanas que actualmente está en la celda D4.

En la actualidad las hojas de cálculo proporcionan tres métodos para realizar análisis de sensibilidad. Uno es verificar el efecto de un cambio individual en el modelo mediante la simple modificación de valores en la hoja de cálculo para después resolver de nuevo. El segundo consiste en generar de manera sistemática una tabla en una sola hoja de cálculo que muestra el efecto de una serie de cambios en uno o dos parámetros del modelo. En el tercero se obtiene y se aplica un informe de sensibilidad con Excel. A continuación se describe cada uno de estos métodos de manera independiente.

Verificación de cambios individuales en el modelo

Una de las grandes fortalezas de una hoja de cálculo es la facilidad con la que se puede usar en forma interactiva para realizar varios tipos de análisis de sensibilidad. Una vez que el Solver se ha preparado para obtener una solución óptima, el usuario puede encontrar de inmediato lo que sucedería si uno de los parámetros del modelo fuera cambiado a algún otro valor. Todo lo que se debe hacer es incorporar este cambio en la hoja de cálculo y después dar clic de nuevo sobre el botón de resolver.

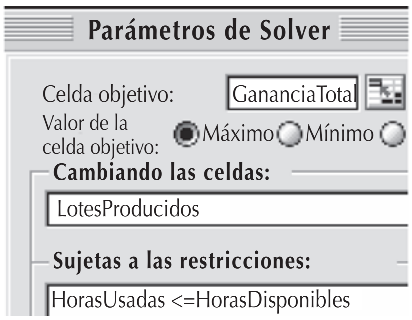
Como ejemplo, suponga que la administración de la Wyndor tiene bastantes dudas acerca de cuál será la ganancia por lote de puertas (G_P). A pesar de que la cifra de \$3 000 dada en la figura 6.8 se considera una estimación inicial razonable, la administración considera que la ganancia real podría desviarse en forma sustancial de esta cifra en cualquier dirección. Sin embargo, el intervalo entre $G_P =$ \$2 000 y $G_P =$ \$5 000 se considera muy confiable.

En la figura 6.9 se muestra lo que pasaría si la ganancia por lote de puertas se redujera de G_P = \$3 000 a G_P = \$2 000. En comparación con la figura 6.8, no hay ningún cambio en la solución óptima de la mezcla de productos. De hecho, los *únicos* cambios en la nueva hoja de cálculo son los valores nuevos de G_P en la celda C4 y una disminución de \$2 000 en la ganancia total que se muestra en la celda G12 (porque cada uno de los dos lotes de puertas producidos por semana proporcionan \$1 000 menos de ganancia). Como la solución óptima no cambia, ahora se sabe que la estimación original de G_P = \$3 000 puede ser *muy alta*, sin que esto invalide la solución óptima del modelo.

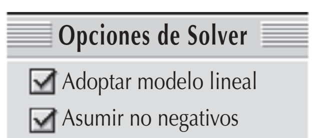
Pero, ¿qué sucede si la estimación inicial de G_P es *muy baja*? En la figura 6.10 se muestra qué pasaría si G_P se incrementara a \$5 000. De nuevo, no hay cambio en la solución óptima. Por tanto, ahora se sabe que el intervalo de valores de G_P sobre el cual la solución óptima actual permanece óptima (es decir, el *rango permisible* para seguir óptima que se estudió en la sección 6.7) incluye al rango que va desde \$2 000 hasta \$5 000 y se puede extender aún más.

Como el valor original de G_P = \$3 000 puede cambiar de manera considerable y en cualquier dirección sin que esto modifique la solución óptima, G_P es un parámetro relativamente insensible. No es necesario hacer esta estimación con gran precisión para tener confianza en que el modelo proporciona la solución óptima correcta.

	A	B	C	D	E	F	G
1	**Problema de mezcla de productos de la Wyndor Glass Co.**						
2							
3			Puertas	Ventanas			
4		Ganancia por lote	$3 000	$5 000			
5					Horas		Horas
6			Horas usadas por lote producido		usadas		disponibles
7		Planta 1	1	0	2	<=	4
8		Planta 2	0	2	12	<=	12
9		Planta 3	3	2	18	<=	18
10							
11			Puertas	Ventanas			Ganancia total
12		Lotes producidos	2	6			$36 000



	E
5	Horas
6	usadas
7	=SUMAPRODUCTO(C7:D7,LotesProducidos)
8	=SUMAPRODUCTO(C8:D8,LotesProducidos)
9	=SUMAPRODUCTO(C9:D9,LotesProducidos)



Nombre de rango	Celdas
LotesProducidos	C12:D12
HorasDisponibles	G7:G9
HorasUsadas	E7:E9
HorasUsadasPorLoteProducido	C7:D9
GananciaPorLote	C4:D4
GananciaTotal	G12

■ **FIGURA 6.8**
Modelo en hoja de cálculo y solución óptima que se obtuvo para el problema original de la Wyndor antes de desarrollar el análisis de sensibilidad.

■ **FIGURA 6.9**
Problema de la Wyndor modificado, donde la estimación de la ganancia por lote de puertas se ha disminuido de $G_P = \$3\,000$ a $G_P = \$2\,000$, pero no hay cambio en la solución óptima para la mezcla de productos.

	A	B	C	D	E	F	G
1	**Problema de mezcla de productos de la Wyndor Glass Co.**						
2							
3			Puertas	Ventanas			
4		Ganancia por lote	$2 000	$5 000			
5					Horas		Horas
6			Horas usadas por lote producido		usadas		disponibles
7		Planta 1	1	0	2	<=	4
8		Planta 2	0	2	12	<=	12
9		Planta 3	3	2	18	<=	18
10							
11			Puertas	Ventanas			Ganancia total
12		Lotes producidos	2	6			$34 000

■ FIGURA 6.10
Problema de la Wyndor modificado, donde la estimación de la ganancia por lote de puertas se ha aumentado $G_P = \$3\ 000$ a $G_P = \$5\ 000$, pero no hay cambio en la solución óptima para la mezcla de productos.

	A	B	C	D	E	F	G
1	**Problema de mezcla de productos de la Wyndor Glass Co.**						
2							
3			Puertas	Ventanas			
4		Ganancia por lote	$5 000	$5 000			
5					Horas		Horas
6			Horas usadas por lote producido		usadas		disponibles
7		Planta 1	1	0	2	<=	4
8		Planta 2	0	2	12	<=	12
9		Planta 3	3	2	18	<=	18
10							
11			Puertas	Ventanas			Ganancia total
12		Ganancia total	2	6			$40 000

Ésta podría ser toda la información que se requiere acerca de G_P. Sin embargo, si existiera una posibilidad grande de que el valor real de G_P pudiera resultar fuera de este amplio rango de \$2 000 a \$5 000, sería deseable una investigación más profunda. ¿Qué tan alto o bajo puede ser G_P antes de que la solución óptima cambie?

En la figura 6.11 se demuestra que la solución óptima cambia si G_P se incrementa hasta $G_P = \$10\ 000$. Así, se sabe que este cambio ocurre en algún punto entre \$5 000 y \$10 000 durante el proceso de incremento de G_P.

Utilización de la Tabla de Solver para hacer análisis de sensibilidad en forma sistemática

Para determinar el punto exacto en el que cambia la solución óptima, se podrían seleccionar más valores en forma aleatoria hasta encontrarlo. Sin embargo, un mejor enfoque es el de considerar de manera sistemática un intervalo de valores de G_P. Un complemento de Excel desarrollado por el profesor Mark Hillier, llamado *Tabla de Solver*, está diseñado para realizar exactamente este tipo de análisis. Se encuentra disponible en el OR Courseware del sitio en internet de este libro. Para instalarlo sólo es necesario abrir el archivo de la Tabla de Solver en el OR Courseware.

La Tabla de Solver se usa para mostrar los resultados en las celdas cambiantes o ciertas celdas de salida de varios valores de prueba en una celda de datos. Para cada valor de prueba de la celda de datos, se resuelve el modelo por medio del Solver. Por tanto, la Tabla de Solver (o cualquier complemento de Excel comparable) proporciona una forma sistemática de realizar análisis de sensibilidad para después desplegar los resultados a los administradores y otros que no estén familiarizados con los aspectos más técnicos del análisis de sensibilidad.

Para usar la Tabla de Solver, en primera instancia expanda la hoja de cálculo original (figura 6.8) para hacer una tabla con encabezados, como se muestra en la figura 6.12. En la primera co-

■ FIGURA 6.11
Problema de la Wyndor modificado, donde la estimación de la ganancia por lote de puertas se ha incrementado de $G_P = \$3\ 000$ a $G_P = \$10\ 000$, lo que genera un cambio de la solución óptima de la mezcla de productos.

	A	B	C	D	E	F	G
1	**Problema de mezcla de productos de la Wyndor Glass Co.**						
2							
3			Puertas	Ventanas			
4		Ganancia por lote	$10 000	$5 000			
5					Horas		Horas
6			Horas usadas por lote producido		usadas		disponibles
7		Planta 1	1	0	4	<=	4
8		Planta 2	0	2	6	<=	12
9		Planta 3	3	2	18	<=	18
10							
11			Puertas	Ventanas			Ganancia total
12		Lotes producidos	4	3			$55 000

lumna de la tabla (celdas B19:B28), liste los valores de prueba de la celda de datos (la ganancia por lote de puertas), pero deje el primer renglón (celda B18) en blanco. Los encabezados de las columnas siguientes especifican qué salida se evaluará. En cada una de dichas columnas, use el primer renglón de la tabla (celdas C18:E18) para escribir una ecuación que fije el valor en cada una de estas celdas a la celda con cambio significativo o la celda de salida. En este caso, las celdas de interés son DoorBatchesProduced (C12), WindowBatchesProduced (D12) y TotalProfit (G12), por lo que las ecuaciones de C18:E18 son las que se muestran justamente debajo de la hoja de cálculo de la figura 6.12.

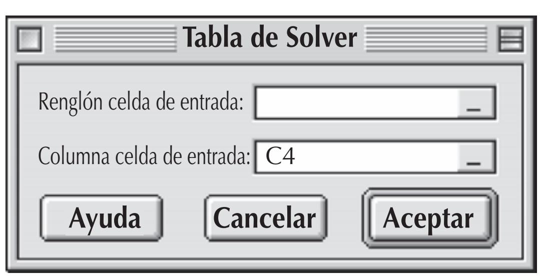
A continuación, seleccione la tabla completa al hacer clic y arrastrar desde la celda B18 hasta la celda E28, posteriormente elija Tabla de Solver de la pestaña Add-Ins (para el caso de Excel

FIGURA 6.12
Expansión de la hoja de cálculo de la figura 6.8 para preparar la utilización de la Tabla de Solver, con la cual se muestra el efecto de la variación sistemática en la estimación de la ganancia por lote de puertas.

	A	B	C	D	E	F	G
1	**Problema de mezcla de productos de la Wyndor Glass Co.**						
2							
3			Puertas	Ventanas			
4		Ganancia por lote	$3 000	$5 000			
5					Horas		Horas
6			Horas usadas por lote producido		usadas		disponibles
7		Planta 1	1	0	2	<=	4
8		Planta 2	0	2	12	<=	12
9		Planta 3	3	2	18	<=	18
10							
11			Puertas	Ventanas			Ganancia total
12		Lotes producidos	2	6			$36 000
13							
14							
15							
16		Ganancia por lote	Lotes producidos óptimos		Ganancia		
17		de puertas	Puertas	Ventanas	total		
18			2	6	$36 000		
19		$1 000					
20		$2 000					
21		$3 000					
22		$4 000					
23		$5 000					
24		$6 000					
25		$7 000					
26		$8 000					
27		$9 000					
28		$10 000					

Seleccione estas celdas (B18:E28) antes de elegir la Tabla de Solver

	C	D	E
16	Lotes producidos óptimos		Ganancia
17	Puertas	Ventanas	Total
18	=LotesDePuertasProducidos	=LotesDeVentanasProducidos	=GananciaTotal



Nombre de rango	Celdas
LotesDePuertasProducidos	C12
GananciaTotal	G12
LotesDeVentanasProducidos	D12

2007) o del menú Herramientas (para el caso de versiones anteriores de Excel), después de haber instalado este complemento de Excel que se proporciona en el OR Courseware. En el cuadro de diálogo de la Tabla de Solver (como se muestra en la parte baja de la figura 6.12), indique la celda de entrada de columna (C4), que se refiere a la celda de datos que cambia en la primera columna de la tabla. No se introduce nada a la celda de entrada de renglón porque en este caso no se usa ningún renglón para enlistar los valores de prueba de una celda de datos.

La Tabla de Solver que se muestra en la figura 6.13 se genera en forma automática al hacer clic sobre el botón de aceptar. Para cada valor de prueba enlistado en la primera columna de la tabla para la celda de datos bajo consideración, Excel resuelve de nuevo el problema mediante el Solver y después llena los valores correspondientes en las otras columnas de la tabla. (Los números del primer renglón de la tabla provienen de la solución original en la hoja de cálculo, antes de que el valor original de la celda de datos haya sido cambiado.)

La tabla revela que la solución óptima permanece igual desde G_P = $1 000 (y quizá más bajo) hasta G_P = $7 000, pero ocurre un cambio en algún punto entre $7 000 y $8 000. En seguida se podrían considerar valores para G_P de manera sistemática entre $7 000 y $8 000 para determinar con mayor exactitud dónde cambia la solución óptima. Sin embargo, esto no es necesario puesto que, como se expuso un poco antes, el uso del informe de sensibilidad de Excel es un atajo para determinar con precisión dónde cambia la solución óptima.

Hasta ahora se ha ilustrado cómo investigar de manera sistemática el efecto de cambiar sólo G_P (celda C4 en la figura 6.8). El enfoque es el mismo en el caso de G_V (celda D4). En realidad, la Tabla de Solver se puede usar de esta forma para investigar el efecto de cambiar *cualquier* celda de datos en el modelo, incluso cualquier celda en HorasDisponibles (G7:G9) u HorasUsadasPorLoteProducido (C7:D9).

A continuación se presentará cómo investigar cambios simultáneos en dos celdas de datos con una hoja de cálculo, primero por sí sola y después con la ayuda de la Tabla de Solver.

FIGURA 6.13
Una aplicación de la Tabla de Solver que muestra el efecto de la variación sistemática en la estimación de la ganancia por lote de puertas en el problema de la Wyndor.

	A	B	C	D	E	F	G
1	**Problema de mezcla de productos de la Wyndor Glass Co.**						
2							
3			Puertas	Ventanas			
4		Ganancia por lote	$3 000	$5 000			
5					Horas		Horas
6			Horas usadas por lote producido		usadas		disponibles
7		Planta 1	1	0	2	<=	4
8		Planta 2	0	2	12	<=	12
9		Planta 3	3	2	18	<=	18
10							
11			Puertas	Ventanas			Ganancia total
12		Lotes producidos	2	6			$36 000
13							
14							
15							
16		Ganancia por lote	Lotes producidos óptimos		Ganancia		
17		de puertas	Puertas	Ventanas	total		
18			2	6	$36 000		
19		$1 000	2	6	$32 000		
20		$2 000	2	6	$34 000		
21		$3 000	2	6	$36 000		
22		$4 000	2	6	$38 000		
23		$5 000	2	6	$40 000		
24		$6 000	2	6	$42 000		
25		$7 000	2	6	$44 000		
26		$8 000	4	3	$47 000		
27		$9 000	4	3	$51 000		
28		$10 000	4	3	$55 000		

Verificación de cambios en el modelo por dos vías

Cuando se usan las estimaciones originales de G_P ($3 000) y G_V ($5 000), la solución óptima que indica el modelo (figura 6.8) está muy inclinada a producir ventanas (seis lotes por semana) en comparación con las puertas (sólo dos lotes por semana). Suponga que la administración de Wyndor está preocupada por este desbalance y siente que el problema puede ser que la estimación de G_P sea demasiado baja y la de G_V sea muy alta. De aquí surge la siguiente pregunta: Si en realidad las estimaciones están equivocadas en las direcciones que se plantean, ¿ello conduciría a una mezcla de productos más balanceada que fuera más provechosa? (Recuerde que la *razón* de G_P a G_V es relevante para determinar la mezcla de productos óptima, de forma que si las estimaciones son erróneas con la realización de un cambio pequeño en la *misma* dirección sería poco probable que la mezcla de productos óptima se modifique.)

Esta pregunta se puede responder en cuestión de segundos sólo con sustituir las nuevas estimaciones de ganancia por lote en la hoja de cálculo original en la figura 6.8 y al hacer clic sobre el botón de resolver. En la figura 6.14 se muestra que las nuevas estimaciones de $4 500 para puertas y $4 000 para ventanas no ocasionan ningún cambio en la mezcla de productos óptima. (La ganancia total sí cambia, pero ello ocurre sólo por las modificaciones en las ganancias por lote.) ¿Podrían cambios aún más grandes en las estimaciones de las ganancias finalmente conducir a un cambio en la mezcla de productos óptima? En la figura 6.15 se muestra que esto sucede, lo que ocasiona una mezcla de productos relativamente balanceada de $(x_1, x_2) = (4, 3)$, cuando se usan las estimaciones de $6 000 para puertas y $3 000 para ventanas.

Las figuras 6.14 y 6.15 no revelan dónde cambia la mezcla de productos óptima a medida que las estimaciones de las ganancias se incrementan de $4 500 a $6 000 en el caso de las puertas y disminuyen de $4 000 a $3 000 en el de las ventanas. A continuación se describe la forma en que la Tabla de Solver puede ayudar en forma sistemática a determinar esta cuestión de mejor modo.

Utilización de la Tabla de Solver para análisis de sensibilidad de dos vías

Una versión de la Tabla de Solver de dos vías proporciona una manera de investigar de manera sistemática el efecto si las estimaciones introducidas en *dos* celdas de datos no son exactas en forma simultánea. (Sin embargo, dos es el número máximo de celdas de datos que se pueden considerar al mismo tiempo en la Tabla de Solver.) En este caso, la Tabla de Solver presenta los resultados en una sola celda de salida para varios valores de prueba en dos celdas de datos.

Para ilustrar este enfoque, de nuevo se investigará el efecto de incrementar G_P y disminuir G_V de manera simultánea. Antes de considerar el efecto sobre la mezcla de productos óptima, se observará cómo afecta la utilidad total. En este caso se usará la Tabla de Solver para mostrar la forma en que GananciaTotal (G12) de la figura 6.8 varía sobre un intervalo de valores de prueba en las dos celdas de datos, GananciaPorLote (C4:D4). Para cada par de valores de prueba de estas dos celdas de datos, se correrá el Solver para que él resuelva de nuevo el problema.

La Tabla de Solver de dos vías para el problema de la Wyndor se crea de la siguiente manera: expanda la hoja de cálculo original (figura 6.8) para hacer una tabla con renglones y encabezados

■ **FIGURA 6.14**
Problema de la Wyndor modificado, donde las estimaciones de las ganancias por lote de puertas y ventanas se han modificado a G_P = $4 500 y G_V = $4 000, respectivamente, pero no hay cambio en la solución óptima de la mezcla de productos.

	A	B	C	D	E	F	G
1	**Problema de mezcla de productos para la Wyndor Glass Co.**						
2							
3			Puertas	Ventanas			
4		Ganancia por lote	$4 500	$4 000			
5					Horas		Horas
6			Horas usadas por lote producido		usadas		disponibles
7		Planta 1	1	0	2	<=	4
8		Planta 2	0	2	12	<=	12
9		Planta 3	3	2	18	<=	18
10							
11			Puertas	Ventanas			Ganancia total
12		Lotes producidos	2	6			$33 000

■ **FIGURA 6.15**
Problema de la Wyndor modificado, donde las estimaciones de las ganancias por lote de puertas y ventanas se han modificado a $G_P = \$6\,000$ a $G_V = \$3\,000$, respectivamente, lo que genera un cambio de la solución óptima de la mezcla de productos.

	A	B	C	D	E	F	G
1	**Problema de mezcla de productos de la Wyndor Glass Co.**						
2							
3			Puertas	Ventanas			
4		Ganancia por lote	$6 000	$3 000			
5					Horas		Horas
6			Horas usadas por lote producido		usadas		disponibles
7		Planta 1	1	0	4	<=	4
8		Planta 2	0	2	6	<=	12
9		Planta 3	3	2	18	<=	18
10							
11			Puertas	Ventanas			Ganancia total
12		Lotes producidos	4	3			$33 000

de columna como los que se muestran en los renglones 16-21 de la hoja de cálculo de la figura 6.16. En la esquina superior izquierda de la tabla (C17), escriba una ecuación (=GananciaTotal) que haga referencia a la celda de interés. En la primera columna de la tabla (columna C, bajo la ecuación de la celda C17), inserte varios valores de prueba para la primera celda de datos en estudio (la ganancia por lote de puertas). En el primer renglón de la tabla (renglón 17, a la derecha de la ecuación de la celda C17), inserte varios valores de prueba para la segunda celda de datos de interés (la ganancia por lote de ventanas).

Luego, seleccione la tabla completa (C17:H21) y luego la Tabla de Solver de la pestaña Add-Ins (en el caso de Excel 2007) o Menú herramientas (para el caso de versiones anteriores de Excel), después de haber instalado este *add-in* de Excel que viene en su OR Courseware. En el cuadro de diálogo de la Tabla de Solver (que se muestra en la parte baja de la figura 6.16) indique cuáles son

■ **FIGURA 6.16**
Expansión de la hoja de cálculo de la figura 6.8 para preparar la utilización de la Tabla de Solver de dos dimensiones, con la cual se muestra el efecto sobre las ganancias totales de la variación sistemática de las estimaciones de las ganancias por lote de puertas y ventanas.

	A	B	C	D	E	F	G	H	I
1	**Problema de mezcla de productos de la Wyndor Glass Co.**								
2									
3			Puertas	Ventanas					
4		Ganancia por lote	$3 000	$5 000					
5					Horas		Horas		
6			Horas usadas por lote producido		usadas		disponibles		
7		Planta 1	1	0	2	<=	4		
8		Planta 2	0	2	12	<=	12		Seleccione estas celdas (C17:H21) antes de elegir la Tabla de Solver
9		Planta 3	3	2	18	<=	18		
10							Ganancia		
11			Puertas	Ventanas			total		
12		Lotes producidos	2	6			$36 000		
13									
14									
15									
16		**Ganancia total**			Ganancia por lote de ventanas				
17			$36 000	$1 000	$2 000	$3 000	$4 000	$5 000	
18			$3 000						
19		Ganancia por lote	$4 000						
20		de puertas	$5 000						
21			$6 000						



	C
17	=GananciaTotal

Nombre de rango	Celda
GananciaTotal	G12

FIGURA 6.17
Una aplicación de dos dimensiones de la Tabla de Solver, con la cual se muestra el efecto sobre la utilidad total óptima de la variación sistemática en las estimaciones de las ganancias por lote de puertas y ventanas para el problema de la Wyndor.

	B	C	D	E	F	G	H
16	**Ganancia total**			Ganancia por lote para puertas			
17		$36 000	$1 000	$2 000	$3 000	$4 000	$5 000
18		$3 000	$15 000	$18 000	$24 000	$30 000	$36 000
19	Ganancia por lote	$4 000	$19 000	$22 000	$26 000	$32 000	$38 000
20	para ventanas	$5 000	$23 000	$26 000	$29 000	$34 000	$40 000
21		$6 000	$27 000	$30 000	$33 000	$36 000	$42 000

las celdas de datos que se cambiarán de manera simultánea. La columna de datos de entrada C4 hace referencia a las celdas de datos cuyos diferentes valores de prueba se enlistan en la primera columna de la tabla (C18:C21), mientras que el renglón de celdas de entrada D4 se refiere a las celdas de datos cuyos distintos valores de prueba se presentan en el primer renglón de la tabla (D17:H17).

Al hacer clic sobre el botón de aceptar se genera de manera automática la Tabla de Solver que se muestra en la figura 6.17. Para cada par de valores de prueba insertados en las dos celdas de datos, Excel resuelve de nuevo el problema mediante Solver y después obtiene la ganancia total y la coloca en su sitio correspondiente de la tabla. (El número en C17 proviene de la celda objetivo de la hoja de cálculo original antes de que los valores originales de las dos celdas de datos hayan sido cambiados.)

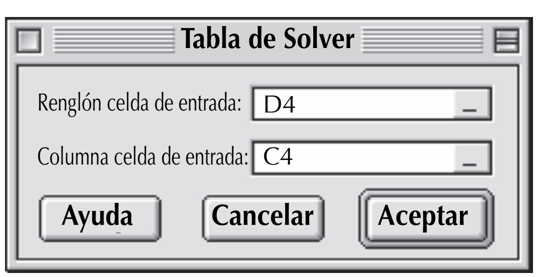
A diferencia de la Tabla de Solver de una vía que puede mostrar los resultados de *múltiples* celdas cambiantes o celdas de salida para varios valores de prueba de *una sola* celda de datos, una Tabla de Solver de dos vías está limitada a mostrar los resultados en una sola celda para cada par de valores de prueba en las dos celdas de datos de interés. Sin embargo, existe un truco en el que se utiliza el símbolo & para permitirle a la Tabla de Solver mostrar los resultados de múltiples celdas cambiantes o celdas de salida dentro de una sola celda de la tabla. En la Tabla Solver que aparece en la figura 6.18 se utiliza este truco para mostrar los resultados de ambas celdas cambiantes, LotesDePuertasProducidos (C12) y LotesDeVentanasProducidos (D12), y de cada par de valores de prueba de GananciaPorLote (C4:D4). La fórmula clave está en la celda C25:

C25 = "("& LotesDePuertasProducidos &", "& LotesDeVentanasProducidos &")"

FIGURA 6.18
Aplicación de dos dimensiones de la Tabla de Solver, con la cual se muestra el efecto sobre la mezcla de productos óptima de la variación sistemática en las estimaciones de las ganancias por lote de puertas y ventanas en el problema de la Wyndor.

	B	C	D	E	F	G	H
24	Lotes producidos (puertas, ventanas)			Ganancia por lote para ventanas			
25		(2,6)	$1 000	$2 000	$3 000	$4 000	$5 000
26		$3 000	(4,3)	(4,3)	(2,6)	(2,6)	(2,6)
27	Ganancia por lote	$4 000	(4,3)	(4,3)	(2,6)	(2,6)	(2,6)
28	para puertas	$5 000	(4,3)	(4,3)	(4,3)	(2,6)	(2,6)
29		$6 000	(4,3)	(4,3)	(4,3)	(4,3)	(4,3)

	C
25	="(" & LotesDePuertasProducidos &", "& LotesDeVentanasProducidos & ")"



Nombre de rango	Celdas
LotesDePuertasProducidos	C12
LotesDeVentanasProducidos	D12

El carácter & le indica a Excel concatenar, de forma que el resultado será un paréntesis izquierdo, seguido por el valor en LotesDePuertasProducidos (C12), después una coma y luego el contenido de LotesDeVentanasProducidos (D12), y finalmente un paréntesis derecho. Si LotesDePuertasProducidos = 2 y LotesDeVentanasProducidos = 6, el resultado es (2, 6). De esta forma, los resultados de *ambas* celdas cambiantes se despliegan en una *sola* celda de la tabla.

Después de los preliminares usuales al introducir la información que se presenta en los renglones 24-25 y las columnas B-C de la figura 6.18, junto con la fórmula en C25, al hacer clic sobre el botón de aceptar se genera de manera automática la Tabla de Solver completa. Las celdas D26: H29 muestran la solución óptima para las diferentes combinaciones de los valores de prueba de las ganancias por lote de puertas y ventanas. La esquina superior derecha (celda H26) de esta Tabla de Solver proporciona la solución óptima de $(x_1, x_2) = (2, 6)$ cuando se utilizan las estimaciones originales de $3 000 por lote de puertas y $5 000 por lote de ventanas. El movimiento hacia abajo a partir de esta celda corresponde a incrementar la estimación para las puertas, mientras que desplazarse hacia la izquierda corresponde a disminuir la estimación de las ventanas. (Las celdas hacia arriba o a la derecha de H26 no se muestran porque estos cambios sólo incrementarían la atracción hacia (2, 6) como la solución óptima.) Observe que $(x_1, x_2) = (2, 6)$ continúa como la solución óptima para todas las celdas cercanas a H26. Esto indica que las estimaciones originales de ganancia por lote deberían ser muy inexactos para que la mezcla óptima de productos pueda cambiar.

Empleo del informe de sensibilidad para realizar análisis de sensibilidad

Hasta ahora se ha visto cómo pueden realizarse algunos análisis de sensibilidad con rapidez en una hoja de cálculo, ya sea al hacer cambios interactivos en las celdas de datos para después resolver una y otra vez o mediante el uso de la Tabla de Solver para generar información similar de manera sistemática. Sin embargo, existe un atajo. Una parte de esta misma información (y más) puede obtenerse con mayor rapidez y precisión con la simple utilización del informe de sensibilidad que proporciona el Solver de Excel. (En esencia, el mismo informe de sensibilidad es una parte estándar de las salidas disponibles en otros paquetes de software de programación lineal, como MPL/CPLEX, LINDO y LINGO.)

Ya se estudió en la sección 4.7 el informe de sensibilidad y cómo se usa para realizar el análisis de sensibilidad. En la figura 4.10 de esa sección se muestra el informe de sensibilidad del problema de la Wyndor. Una parte de este informe se muestra en la figura 6.19. En lugar de repetir la sección 4.7, ahora se dará prioridad a la ilustración de la forma en que el informe de sensibilidad puede contestar de manera eficiente las preguntas específicas surgidas en las subsecciones anteriores del problema de la Wyndor.

La pregunta considerada en las primeras dos subsecciones fue: ¿cuán lejos puede llevarse la estimación inicial de $3 000 de G_P sin que la solución óptima actual, $(x_1, x_2) = (2, 6)$, cambie? En las figuras 6.10 y 6.11 se mostró que la solución óptima no cambiaría hasta que G_P se elevara a algún valor entre $5 000 y $10 000. En la figura 6.13 se redujo el intervalo en el que cambia la solución óptima a algún punto entre $7 000 y $8 000. En esta figura también se mostró que si la estimación inicial de $3 000 para G_P es muy alto en lugar de ser muy bajo, G_P necesitaría reducirse hasta por debajo de $1 000 antes de que la solución óptima pudiera cambiar.

Ahora observe que la parte del informe de sensibilidad de la figura 6.19 aborda esta misma pregunta. El renglón de LotesDePuertasProducidos en este informe proporciona la siguiente información acerca de G_P (sin los signos de dólar).

Valor actual de G_P:	3 000.	
Aumento permisible en G_P:	4 500.	entonces $G_P \leq 3\,000 + 4\,500 = 7\,500$
Aumento permisible en G_P:	3 000.	entonces $G_P \geq 3\,000 - 3\,000 = 0$.
Rango permisible para G_P:		$0 \leq G_P \leq 7\,500$.

Por tanto, si G_P cambia de su valor actual (sin hacer ningún otro cambio en el modelo), la solución actual $(x_1, x_2) = (2, 6)$ permanecerá óptima mientras G_P esté dentro de este *rango permisible* para seguir óptima, $0 \leq G_P \leq \$7\,500$.

En la figura 6.20 se proporciona una visión gráfica de este rango permisible para seguir óptima. En el caso del valor original de $G_P = 3\,000$, la línea continua de la figura muestra la pendiente de la

■ **FIGURA 6.19**
Parte del informe de sensibilidad generado por Excel Solver para el problema original de la Wyndor (figura 6.8), donde las últimas tres columnas identifican los rangos permisibles para las ganancias por lote de puertas y ventanas.

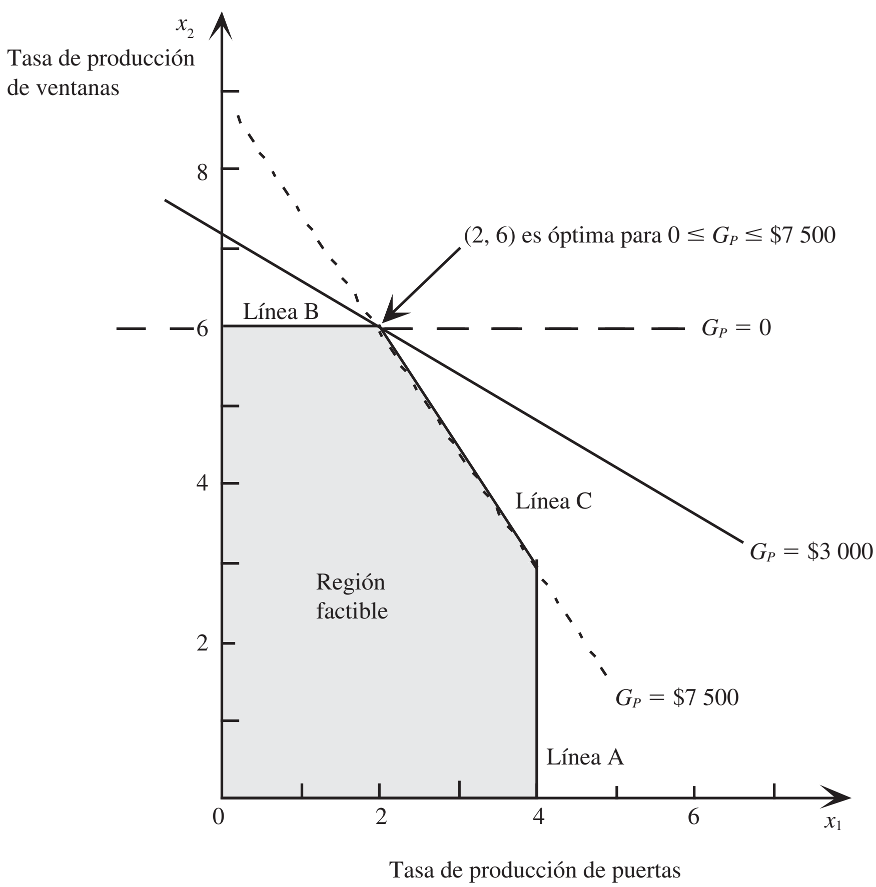
Celdas ajustables

Celda	Nombre	Valor final	Costo reducido	Coeficiente objetivo	Aumento permisible	Disminución permisible
$C$12	LotesDePuertasProducidos	2	0	3 000	4 500	3 000
$D$12	LotesDeVentanasProducidos	6	0	5 000	1E+30	3 000

recta de la función objetivo que pasa por (2, 6). En el extremo inferior del rango permisible para seguir óptima, donde $G_P = 0$, la recta de la función objetivo que pasa por (2, 6) ahora es la línea B de la figura, por lo cual cualquier punto sobre el segmento de recta entre (0, 6) y (2, 6) es una solución óptima. Para cualquier valor de $G_P < 0$, la recta de la función objetivo habrá rotado aún más de forma que (0, 6) se convierte en la única solución óptima. En el extremo superior del rango permisible para seguir óptima, cuando $G_P = 7\,500$, la recta de la función objetivo que pasa por (2, 6) se convierte en la línea C, por lo que cualquier punto sobre el segmento de recta entre (2, 6) y (4, 3) es una solución óptima. Para cualquier valor de $G_P = 7\,500$, la recta de la función objetivo está aún más inclinada que la línea C, por lo cual (4, 3) se convierte en la única solución óptima. En consecuencia, la solución óptima original, $(x_1, x_2) = (2, 6)$, sigue óptima sólo mientras $0 \leq G_P \leq \$7\,500$.

La rutina llamada *Graphical Method and Sensitivity Analysis* (Método gráfico y análisis de sensibilidad) del IOR Tutorial está diseñada para ayudarle a realizar este tipo de análisis gráfico.

■ **FIGURA 6.20**
Las dos líneas punteadas que pasan por las rectas de frontera de restricción sólidas son las rectas de la función objetivo cuando G_P (la ganancia por lote de puertas) está en un extremo de su rango permisible para seguir óptima, $0 \leq G_P$ \$7 500, dado que cualquier recta o función objetivo que esté en medio sigue con $(x_1, x_2) = (2, 6)$ como una solución óptima para el problema de la Wyndor.



Después de introducir el modelo para el problema original de la Wyndor, el módulo proporciona la gráfica que se muestra en la figura 6.20 (sin las líneas punteadas). Sólo es necesario arrastrar un extremo de la recta objetivo hacia arriba o hacia abajo para ver cuánto se puede aumentar o disminuir G_P antes de que $(x_1, x_2) = (2, 6)$ ya no sea óptima.

> *Conclusión:* El rango permisible para que G_P continue siendo óptimo es $0 \leq G_P \leq \$7\,500$, porque $(x_1, x_2) = (2, 6)$ permanece óptima en este intervalo pero no fuera de él. (Cuando $G_P = 0$ o $G_P = \$7\,500$, existen múltiples soluciones, pero $(x_1, x_2) = (2, 6)$ aún es una de ellas.) Con un rango como éste alrededor de la estimación original de \$3 000 ($G_P =$ \$3 000) de la ganancia por lote de puertas, se puede tener la confianza de que se obtendrá la solución óptima correcta de la ganancia real.

Ahora se abordará la pregunta considerada en las dos subsecciones anteriores. ¿Qué pasaría si la estimación de G_P (\$3 000) fuera demasiado baja y la de G_V (\$5 000) demasiado alta al mismo tiempo? En específico, ¿cuánto puede estar equivocada la estimación en estas direcciones antes de que la solución óptima actual, $(x_1, x_2) = (2, 6)$, pueda cambiar?

En la figura 6.14 se mostró que si G_P se aumentara en \$1 500 (de \$3 000 a \$4 500) y G_V disminuyera en \$1 000 (de \$5 000 a \$4 000), la solución óptima permanecería sin cambio. Después, en la figura 6.15 se indicó que si se duplican estos cambios se obtendría un cambio en la solución óptima. Sin embargo, no está claro dónde ocurre este cambio. En la figura 6.18 se proporcionó más información, pero no una respuesta definitiva a esta pregunta.

Por fortuna, a partir del reporte de sensibilidad (figura 6.19), se puede obtener información adicional mediante la utilización de los aumentos y disminuciones permisibles en G_P y G_V. La clave es aplicar la siguiente regla (como ya se estableció en la sección 6.7):

> **Regla de 100% para cambios simultáneos en los coeficientes de la función objetivo:** Si se hacen cambios simultáneos en los coeficientes de la función objetivo, para cada cambio se calcula el porcentaje del cambio permisible (aumento o disminución) para que ese coeficiente siga dentro de su intervalo permisible para continuar óptimo. Si la *suma* de los porcentajes de cambios *no* excede 100%, la solución óptima original definitivamente será todavía óptima. (Si la *suma* excede de 100%, entonces no hay seguridad.)

La regla no especifica qué pasa cuando el porcentaje de cambios *excede* de 100%. La consecuencia depende de las direcciones de los cambios en los coeficientes. Recuerde que son las *razones* de los coeficientes las que determinan la solución óptima, por lo que, en realidad, la solución óptima original permanece óptima aun cuando la suma de los cambios en los porcentajes excede por mucho 100% si los cambios en los coeficientes tienen la misma dirección. Por tanto, exceder de 100% puede o no cambiar la solución óptima, pero mientras no se le sobrepase, la solución óptima original *definitivamente* continuará como tal.

Recuerde que se puede usar con seguridad el intervalo completo de incremento o disminución en un solo coeficiente de la función objetivo sólo si ninguno de los otros coeficientes ha cambiado. Si ocurren cambios simultáneos en los coeficientes, el enfoque debe estar en el *porcentaje* del aumento o disminución permisible que se utiliza para cada coeficiente.

Para ilustrar lo que decimos, considere de nuevo el problema de la Wyndor, junto con la información que proporciona el informe de sensibilidad de la figura 6.19. Suponga ahora que la estimación de G_P se ha incrementado de \$3 000 a \$4 500 mientras que la de G_V ha disminuido de \$5 000 a \$4 000. Los cálculos según la regla de 100% ahora son

G_P: \$3 000 → \$4 500.

$$\text{Porcentaje de aumento permisible} = 100\left(\frac{4\,500 - 3\,000}{4\,500}\right)\% = 33\frac{1}{3}\%$$

G_V: \$5 000 → \$4 000.

$$\text{Porcentaje de disminución permisible} = 100\left(\frac{5\,000 - 4\,000}{3\,000}\right)\% = 33\frac{1}{3}\%$$

$$\text{Suma} = 66\frac{2}{3}\%.$$

Como la suma de los porcentajes no es mayor a 100%, la solución óptima original $(x_1, x_2) = (2, 6)$ definitivamente aún es óptima, tal como se encontró con anterioridad en la figura 6.14.

Ahora suponga que la estimación de G_P se ha incrementado de \$3 000 a \$6 000 mientras que la de G_V ha disminuido de \$5 000 a \$3 000. Los cálculos para la regla de 100% ahora son

G_P: \$3 000 → \$6 000.

$$\text{Porcentaje de aumento permisible} = 100\left(\frac{6\,000 - 3\,000}{4\,500}\right)\% = 66\frac{2}{3}\%$$

G_V: \$5 000 → \$3 000.

$$\text{Porcentaje de disminución permisible} = 100\left(\frac{5\,000 - 3\,000}{3\,000}\right)\% = 66\frac{2}{3}\%$$

$$\text{Suma} = 133\frac{1}{3}\%.$$

Como la suma de los porcentajes ahora supera a 100%, la regla dice que ya no se puede garantizar que $(x_1, x_2) = (2, 6)$ aún sea óptima. De hecho, pronto se descubrirá en las figuras 6.15 y 6.18 que la solución óptima ha cambiado a $(x_1, x_2) = (4, 3)$.

Estos resultados sugieren cómo encontrar el lugar exacto en el que la solución óptima cambia cuando G_P aumenta y G_V disminuye en estas cantidades relativas. Como 100% es el punto medio entre $66\frac{2}{3}$ y $133\frac{1}{3}$, la suma de los porcentajes de cambio será igual a 100% cuando los valores de G_P y G_V estén en los puntos medios entre sus valores de los casos anteriores. En particular, $G_P =$ \$5 250 es el punto medio entre \$4 500 y \$6 000 y $G_V =$ \$3 500 es el punto medio entre \$4 000 y \$3 000. Los cálculos de la regla de 100% ahora son

G_P: \$3 000 → \$5 250.

$$\text{Porcentaje de aumento permisible} = 100\left(\frac{5\,250 - 3\,000}{4\,500}\right)\% = 50\%$$

G_V: \$5 000 → \$3 500.

$$\text{Porcentaje de disminución permisible} = 100\left(\frac{5\,000 - 3\,500}{3\,000}\right)\% = 50\%$$

$$\text{Suma} = 100\%.$$

A pesar de que la suma de los porcentajes es igual a 100%, el hecho de que no lo *exceda* garantiza que $(x_1, x_2) = (2, 6)$ aún es óptima. En la figura 6.21 se muestra de manera gráfica que ahora, *tanto* (2, 6), como (4, 3) son óptimas, así como todos los puntos sobre el segmento de recta que conecta estos dos puntos. Sin embargo, si G_P y G_V se cambiaran por cualquier cantidad más allá de éstas (de forma que la suma de los porcentajes excediera a 100%), la recta de la función objetivo rotaría tanto con respecto a la vertical que $(x_1, x_2) = (4, 3)$ se convertiría en la única solución óptima.

Al mismo tiempo recuerde que si la suma de porcentajes de los cambios permisibles exceden el 100%, la solución óptima no cambia de manera automática. Por ejemplo, suponga que las estimaciones de las ganancias de ambas unidades se dividen entre dos. Los cálculos resultantes de acuerdo con la regla de 100% son

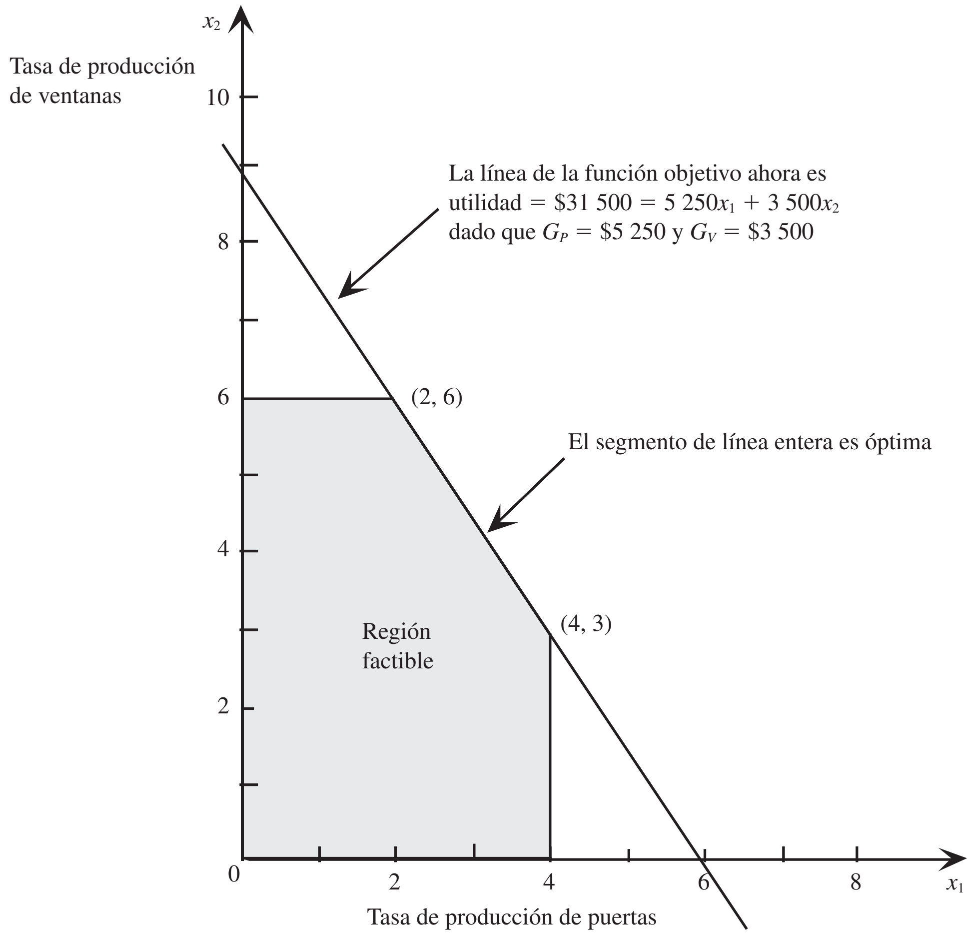
G_P: \$3 000 → \$1 500.

$$\text{Porcentaje de disminución permisible} = 100\left(\frac{3\,000 - 1\,500}{3\,000}\right)\% = 50\%$$

G_V: \$5 000 → \$2 500.

$$\text{Porcentaje de disminución permisible} = 100\left(\frac{5\,000 - 2\,500}{3\,000}\right)\% = 83\frac{1}{3}\%$$

$$\text{Suma} = 133\frac{1}{3}\%.$$



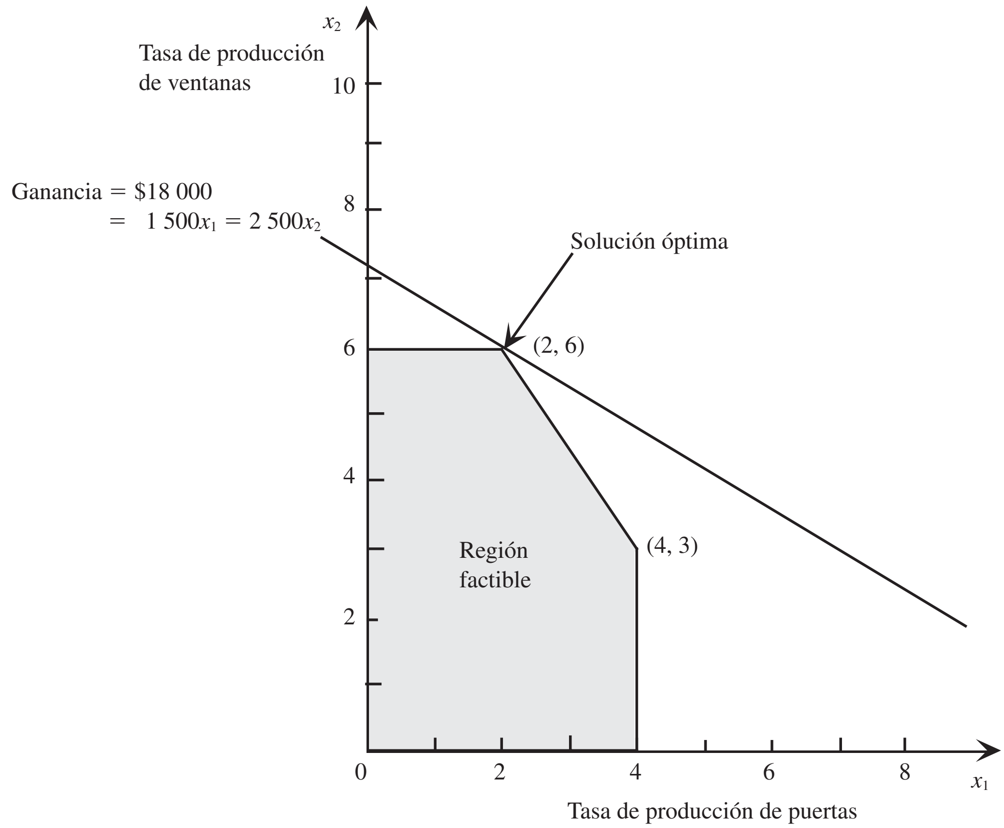
■ **FIGURA 6.21**
Cuando las estimaciones de las ganancias por lote de puertas y ventanas cambian a G_P = \$5 250 y G_V = \$3 500, que están sobre el límite de lo permitido por la regla de 100%, el método gráfico muestra que $(x_1, x_2) = (2, 6)$ todavía es una solución óptima, pero ahora, cualquier punto sobre el segmento de línea entre esta solución y (4, 3) también es óptimo.

Aun cuando esta suma es mayor que 100%, en la figura 6.22 se muestra que la solución óptima original aún es óptima. De hecho, la recta de la función objetivo tiene la misma pendiente que la recta de la función objetivo original (línea continua de la figura 6.20). Esto sucede siempre que se hacen *cambios proporcionales* a todas las estimaciones de ganancias, lo cual de manera automática conduce a la misma solución óptima.

Otros tipos de análisis de sensibilidad

Esta sección se ha enfocado en la forma en que se usa una hoja de cálculo para investigar el efecto de cambios sólo en los coeficientes de las variables de la función objetivo. Con frecuencia también es interesante investigar el efecto de cambios en los lados derechos de las restricciones funcionales. En ocasiones también se puede tener interés en verificar si la solución óptima cambiaría si fuese necesario hacer cambios en algunos coeficientes de las restricciones funcionales.

El enfoque de la hoja de cálculo para investigar estos otros tipos de cambios en el modelo es casi el mismo que en el caso de los coeficientes de la función objetivo. Una vez más, es posible probar algunos cambios en las celdas de datos sólo con hacer estas modificaciones en la hoja de cálculo y utilizar el Solver de Excel para volver a encontrar la solución del modelo. También se puede verificar de manera sistemática el efecto de una serie de cambios en una o dos celdas de datos mediante el uso de la Tabla de Solver. Como ya se describió en la sección 4.7, el informe de sensibilidad generado por el Solver de Excel (o cualquier otro paquete de software de programación lineal), también proporciona información valiosa, inclusive los precios sombra, en relación con el efecto de cambiar el lado derecho de cualquier restricción funcional particular. Cuando se cambia



■ **FIGURA 6.22**
Cuando las estimaciones de las ganancias por lote de puertas y ventanas cambian a $G_P = \$1\,500$ y $G_V = \$2\,500$ (la mitad de sus valores originales), el método gráfico muestra que la solución óptima aún es $(x_1, x_2) = (2, 6)$, aun cuando la regla de 100% dice que la solución óptima podría cambiar.

un cierto número de lados derechos en forma simultánea también existe una "regla de 100%" para este caso, la cual es análoga a la regla de 100% para cambios simultáneos en las restricciones de la función objetivo. (Vea la parte del caso 1 en la sección 6.7 para obtener detalles acerca de cómo investigar el efecto de cambios en los lados derechos, donde se incluye la aplicación de la regla de 100% para cambios simultáneos en los lados derechos.)

En la sección Worked Examples del sitio en internet de este libro se incluyen **otros ejemplos** de la utilización de una hoja de cálculo para investigar el efecto de cambios individuales en los lados derechos.

6.9 CONCLUSIONES

Todo problema de programación lineal tiene, asociado a él, un problema dual de programación lineal. Existen ciertas relaciones útiles entre el problema original (primal) y su problema dual que refuerzan la capacidad para analizar el problema primal. Por ejemplo, la interpretación económica del problema dual proporciona los precios sombra que miden el valor marginal de los recursos del problema primal, al igual que permite interpretar el método símplex. Como este método se puede aplicar en forma directa a cualquiera de los dos problemas para obtener la solución simultánea de ambos, es posible ahorrar una gran cantidad de esfuerzo computacional si se maneja de manera directa el problema dual. La teoría de dualidad, que incluye el método símplex dual (sección 7.1) para trabajar con soluciones básicas superóptimas, tiene un papel muy importante en el análisis de sensibilidad.

Los valores que se usan como parámetros de un modelo de programación lineal son sólo estimaciones. Por tanto, es necesario realizar el análisis de sensibilidad para investigar lo que ocurre si las estimaciones son erróneas. La idea fundamental de la sección 5.3 proporciona la clave para realizar esta investigación de manera eficiente. Los objetivos generales son identificar los parámetros relativamente sensibles que afectan la solución óptima, intentar estimarlos con más cuidado y

después elegir una solución que se mantenga como buena en un cierto intervalo de valores posibles de estos parámetros sensibles. Este análisis constituye una parte muy importante de los estudios de programación lineal.

Las hojas de cálculo, con la ayuda del Solver de Excel, también proporcionan algunos métodos útiles para realizar análisis de sensibilidad. Un método es introducir en forma repetida cambios en uno o más parámetros del modelo en la hoja de cálculo y después hacer clic en el botón de resolver para ver de inmediato si la solución óptima cambia. Un segundo método consiste en utilizar la tabla de Solver para verificar en forma sistemática el efecto de hacer una serie de cambios en uno o dos parámetros del modelo. Un tercero es usar el informe de sensibilidad que proporciona el Solver de Excel para identificar el rango permisible para que los coeficientes mantegan su optimalidad en la función objetivo, los precios sombra de las restricciones funcionales y el rango permisible para cada lado derecho en el que su precio sombra permanece válido. (Otros paquetes que aplican el método símplex —como MPL/CPLEX, LINDO y LINGO— también proporcionan, si así se les solicita, un informe de sensibilidad como éste.)

REFERENCIAS SELECCIONADAS

1. Bertsimas, D. y M. Sim, "The Price of Robustness", en *Operations Research*, **52**(1): 35-53, enero-febrero de 2004.
2. Dantzig, G. B. y M. N. Thapa, *Linear Programming 1: Introduction*, Springer, Nueva York, 1997.
3. Gal, T. y H. Greenberg (eds.), *Advances in Sensitivity Analysis and Parametric Programming*, Kluwer Academic Publishers (actualmente Springer), Boston, MA, 1997.
4. Higle, J. L. y S. W. Wallace, "Sensitivity Analysis and Uncertainty in Linear Programming", en *Interfaces*, **33**(4): 53-60, julio-agosto de 2003.
5. Hillier, F. S. y M. S. Hillier, *Introduction to Management Science: A Modeling and Case Studies Approach with Spreadsheets*, 3a. ed., McGraw-Hill/Irwin, Burr Ridge, IL, 2008, cap. 5.
6. Nazareth, J. L., *An Optimization Primer: On Models, Algorithms and Duality*, Springer-Verlag, Nueva York, 2004.
7. Vanderbei, R. J., *Linear Programming: Foundations and Extensions*, 3a. ed., Springer, Nueva York, 2008.
8. Wendell, R.E., "Tolerance Sensitivity and Optimality Bounds in Linear Programming", en *Management Science*, **50**(6): 797-803, junio de 2004.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos para el capítulo 6

Ejemplo de demostración en OR Tutor:

Análisis de sensibilidad

Rutinas interactivas en IOR Tutorial:

Método gráfico interactivo (Interactive Graphical Method)
Introducción o revisión de un modelo general de programación lineal (Enter or Revise a General Linear Programming Model)
Solución interactiva por el método símplex (Solve Interactively by the Simplex Method)
Análisis de sensibilidad (Sensitivity Analysis)

Rutinas automáticas en IOR Tutorial:

Solución automática por el método símplex (Solve Automatically by the Simplex Method)
Método gráfico y análisis de sensibilidad (Graphical Method and Sensitivity Analysis)

Complementos de Excel:

Premium Solver for Education
Solver Table

Archivos (capítulo 3) para resolver el ejemplo de la Wyndor:

Archivos de Excel
Archivo LINGO/LINDO
Archivo MPL/CPLEX

Glosario del capítulo 6

Vea el apéndice 1 para la documentación del software.

PROBLEMAS

Los símbolos a la izquierda de algunos problemas (o de sus incisos) significan lo siguiente:

D: El ejemplo de demostración indicado puede ser útil.
I: Se sugiere que use las rutinas interactivas indicadas (la impresión registra su trabajo).
C: Use alguna opción de software disponible (o el que indique su profesor) para la solución automática del problema.
E*: Utilice Excel.

Un asterisco en el número del problema indica que al final del libro se proporciona al menos una respuesta parcial.

6.1-1.* Construya la tabla primal-dual y el problema dual para cada uno de los siguientes modelos de programación lineal que se ajustan a nuestra forma estándar.
a) El modelo del problema 3.1-6
b) El modelo del problema 4.7-5

6.1-2. Considere el modelo de programación lineal que aparece en el problema 4.5-4.
a) Construya la tabla primal-dual y el problema dual de este modelo.
b) ¿Qué implica para el problema dual el hecho de que Z no esté acotada en este modelo?

6.1-3. En el caso de cada uno de los siguientes modelos de programación lineal, proporcione su recomendación sobre la manera (tal vez) más eficiente de obtener una solución: aplicación del método símplex directamente a este problema primal o al problema dual. Justifique sus respuestas.
a) Maximizar $\quad Z = 10x_1 - 4x_2 + 7x_3,$

sujeta a

$$\begin{aligned} 3x_1 - x_2 + 2x_3 &\leq 25 \\ x_1 - 2x_2 + 3x_3 &\leq 25 \\ 5x_1 + x_2 + 2x_3 &\leq 40 \\ x_1 + x_2 + x_3 &\leq 90 \\ 2x_1 - x_2 + x_3 &\leq 20 \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

b) Maximizar $\quad Z = 2x_1 + 5x_2 + 3x_3 + 4x_4 + x_5,$

sujeta a

$$\begin{aligned} x_1 + 3x_2 + 2x_3 + 3x_4 + x_5 &\leq 6 \\ 4x_1 + 6x_2 + 5x_3 + 7x_4 + x_5 &\leq 15 \end{aligned}$$

y

$$x_j \geq 0, \quad \text{para } j = 1, 2, 3, 4, 5.$$

6.1-4. Considere el siguiente problema.

Maximizar $\quad Z = -x_1 - 2x_2 - x_3,$

sujeta a

$$\begin{aligned} x_1 + x_2 + 2x_3 &\leq 12 \\ x_1 + x_2 - x_3 &\leq 1 \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

a) Construya el problema dual.
b) Use la teoría de la dualidad para demostrar que la solución óptima del problema primal tiene $Z \leq 0$.

6.1-5. Considere el siguiente problema.

Maximizar $\quad Z = 5x_1 + 4x_2 + 3x_3,$

sujeta a

$$\begin{aligned} x_1 + x_3 &\leq 15 \quad \text{(recurso 1)} \\ x_2 + 2x_3 &\leq 25 \quad \text{(recurso 2)} \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

a) Construya el problema dual de este problema primal.
I ***b***) Resuelva el problema dual en forma gráfica. Utilice esta solución para identificar los precios sombra de los recursos del problema primal.
C ***c***) Confirme sus resultados del inciso *b*) al resolver el problema primal en forma automática mediante el método símplex; después, identifique los precios sombra.

6.1-6. Siga las instrucciones del problema 6.1-5 para resolver el siguiente problema.

Maximizar $Z = x_1 - 3x_2 + 2x_3$,

sujeta a

$$2x_1 + 2x_2 - 2x_3 \leq 6 \quad \text{(recurso 1)}$$
$$-x_2 + 2x_3 \leq 4 \quad \text{(recurso 2)}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

6.1-7. Considere el siguiente problema.

Maximizar $Z = 2x_1 + 3x_2$,

sujeta a

$$4x_1 + x_2 \leq 20$$
$$-x_1 + x_2 \leq 10$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

I ***a***) Demuestre en forma gráfica que este problema no tiene soluciones factibles.
b) Construya el problema dual.
I ***c***) Demuestre en forma gráfica que el problema dual tiene una función objetivo no acotada.

I **6.1-8.** Construya y grafique un problema primal con dos variables de decisión y dos restricciones funcionales que tenga soluciones factibles y una función objetivo no acotada. Después, construya el problema dual y demuestre en forma gráfica que no tiene soluciones factibles.

I **6.1-9.** Construya un par de problemas primal y dual, cada uno con dos variables de decisión y dos restricciones funcionales, tales que ninguno tenga soluciones factibles. Demuestre esta propiedad en forma gráfica.

6.1-10. Construya un par de problemas primal y dual, cada uno con dos variables de decisión y dos restricciones funcionales, tales que el primal no tenga soluciones factibles y que el dual tenga una función objetivo no acotada.

6.1-11. Utilice la propiedad de dualidad débil para demostrar que si ambos problemas, el primal y el dual, tienen soluciones factibles, entonces ambos deben tener una solución óptima.

6.1-12. Considere los problemas primal y dual en la forma estándar que se presentó en la notación matricial al principio de la sección 6.1. Utilice sólo esta definición del problema dual de un problema primal en esta forma para demostrar cada uno de los siguientes resultados.
a) La propiedad de dualidad débil que se presentó en la sección 6.1.
b) Si el problema primal tiene una región factible no acotada que permite aumentar Z de manera indefinida, entonces el problema dual no tiene soluciones factibles.

6.1-13. Considere los problemas primal y dual en nuestra forma estándar presentados en notación matricial al principio de la sección 6.1. Sea $\mathbf{y}^*$ la solución óptima de este problema dual. Suponga que $\mathbf{b}$ se sustituye por $\overline{\mathbf{b}}$. Sea $\overline{x}$ la solución óptima del nuevo problema primal. Demuestre que

$$\mathbf{c}\overline{\mathbf{x}} \leq \mathbf{y}^*\overline{\mathbf{b}}.$$

6.1-14. Para cualquier problema de programación lineal en nuestra forma estándar y su problema dual, diga si las siguientes afirmaciones son falsas o verdaderas y justifique su respuesta.
a) La suma del número de restricciones funcionales y del número de variables (antes de aumentar el problema) es la misma en ambos problemas, primal y dual.
b) En cada iteración, el método símplex identifica al mismo tiempo una solución FEV para el problema primal y una solución FEV para el problema dual tales que los valores de su función objetivo son iguales.
c) Si el problema primal tiene una función objetivo no acotada, entonces el valor óptimo de la función objetivo del problema dual debe ser igual a cero.

6.2-1. Considere la tabla símplex del problema de la Wyndor Glass Co., que se presentó en la tabla 4.8. En el caso de cada tabla símplex, dé la interpretación económica de los siguientes elementos:
a) Cada uno de los coeficientes de las variables de holgura (x_3, x_4, x_5) del renglón 0
b) Cada uno de los coeficientes de las variables de decisión (x_1, x_2) del renglón 0
c) Las alternativas que resultan para la variable básica entrante (o la decisión de detenerse después de la tabla símplex final)

6.3-1.* Considere el siguiente problema.

Maximizar $Z = 6x_1 + 8x_2$,

sujeta a

$$5x_1 + 2x_2 \leq 20$$
$$x_1 + 2x_2 \leq 10$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

a) Construya el problema dual de este problema primal.
b) Resuelva ambos problemas en forma gráfica. Identifique las soluciones FEV y las soluciones no factibles en un vértice para ambos problemas. Calcule los valores de la función objetivo de todas las soluciones.
c) Con la información que obtuvo en el inciso *b*) construya una tabla que presente las soluciones básicas complementarias de estos problemas. (Use los encabezados de la tabla 6.9.)
I ***d***) Trabaje el método símplex paso a paso para resolver el problema. Después de cada iteración (inclusive la 0), identifique las soluciones BF para este problema y las básicas complementarias para el dual. También identifique las soluciones en los vértices correspondientes.

6.3-2. Considere el modelo con dos restricciones funcionales y dos variables que se presentó en el problema 4.1-5. Siga las instrucciones del problema 6.3-1 para este modelo.

6.3-3. Considere los problemas primal y dual del ejemplo de la Wyndor Glass Co., que se presentaron en la tabla 6.1. Use las tablas 5.5, 5.6, 6.8 y 6.9 para construir una nueva tabla que incluya los ocho conjuntos de variables no básicas del problema primal de la columna 1,

los conjuntos correspondientes de variables asociadas del problema dual de la columna 2 y el conjunto de variables no básicas de cada solución básica complementaria del problema dual de la columna 3. Explique por qué esta tabla demuestra la propiedad de holgura complementaria de este ejemplo.

6.3-4. Suponga que el problema primal tiene una solución BF *degenerada* (una o más variables básicas iguales a cero) como su solución óptima. ¿Qué ocasiona en el problema dual esta degeneración? ¿Por qué? ¿La situación inversa también es cierta?

6.3-5. Considere el siguiente problema.

Maximizar $Z = 3x_1 - 8x_2$,

sujeta a

$$x_1 - 2x_2 \leq 10$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

a) Construya el problema dual y después encuentre su solución óptima por inspección.
b) Utilice la propiedad de holgura complementaria y la solución óptima del problema dual para encontrar la solución óptima del problema primal.
c) Suponga que c_1, el coeficiente de x_1 en la función objetivo del primal, puede tener cualquier valor en el modelo. ¿Para cuáles valores de c_1 ocurre que el problema dual no tiene soluciones factibles? En el caso de estos valores, ¿qué es lo que la teoría de dualidad implica sobre el problema primal?

6.3-6. Considere el siguiente problema.

Maximizar $Z = 2x_1 + 7x_2 + 4x_3$,

sujeta a

$$x_1 + 2x_2 + x_3 \leq 10$$
$$3x_1 + 3x_2 + 2x_3 \leq 10$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0.$$

a) Construya el problema dual de este problema primal.
b) Utilice el problema dual para demostrar que los valores óptimos de *Z* en el primal no pueden exceder de 25.
c) Se ha conjeturado que x_2 y x_3 deben ser variables básicas de la solución óptima del problema primal. Obtenga directamente esta solución óptima (y *Z*) mediante el método de eliminación gaussiana. Al mismo tiempo, obtenga e identifique la solución básica complementaria del problema dual por medio de la ecuación (0) del primal. Después, explique si estas dos soluciones son óptimas para sus respectivos problemas.
I ***d***) Resuelva el problema dual en forma gráfica. Utilice esta solución para identificar las variables básicas y no básicas de la solución óptima para el primal. Obtenga esta solución con el método de eliminación gaussiana.

6.3-7.* Reconsidere el modelo del problema 6.1-3*b*.
a) Construya su problema dual.
I ***b***) Resuelva este problema dual en forma gráfica.
c) Utilice el resultado de *b*) para identificar las variables básicas y no básicas de la solución óptima del primal.
d) Utilice los resultados de *c*) para obtener la solución básica óptima del problema primal, utilice la eliminación de Gauss para obtener las variables básicas y comience con el sistema inicial de ecuaciones [excluya la ecuación (0)] construido por el método símplex y establezca las variables no básicas iguales a cero.
e) Utilice los resultados del inciso *c*) para identificar las ecuaciones de definición (vea la sección 5.1) de la solución FEV óptima del problema primal, y después use estas ecuaciones para encontrar la solución.

6.3-8. Considere el modelo que se presentó en el problema 5.3-10.
a) Construya el problema dual.
b) Utilice la información que se proporcionó sobre las variables básicas en la solución óptima primal para identificar las variables básicas y no básicas de la solución óptima dual.
c) Utilice los resultados del inciso *b*) para identificar las ecuaciones de definición (vea la sección 5 .1) de la solución FEV óptima correspondiente al problema dual y con estas ecuaciones encuentre la solución.
I ***d***) Resuelva el problema dual en forma gráfica para verificar los resultados del inciso *c*).

6.3-9. Considere el modelo del problema 3.1-5.
a) Construya el problema dual de este modelo.
b) Utilice el hecho de que en el problema primal, $(x_1, x_2) = (13, 5)$ es óptima, para identificar las variables básicas y no básicas de la solución BF óptima del dual.
c) Identifique esta solución óptima del problema dual al obtener en forma directa la ecuación (0) que corresponde a la solución primal óptima identificada en *b*). Obtenga la ecuación por eliminación de Gauss.
d) A partir de los resultados del inciso *b*) identifique las ecuaciones de definición (vea la sección 5.1) de la solución FEV óptima del problema dual. Verifique la solución óptima dual de *c*), y compruebe que satisface este sistema de ecuaciones.

6.3-10. Suponga que también se quiere conocer la información sobre el problema dual al aplicar el método símplex revisado (vea la sección 5.2) al primal en nuestra forma estándar.
a) ¿Cómo se identificaría la solución óptima del dual?
b) Después de obtener la solución BF en cada iteración, ¿cómo se identifica la solución básica complementaria del problema dual?

6.4-1. Considere el siguiente problema.

Maximizar $Z = 5x_1 + 4x_2$,

sujeta a

$$2x_1 + 3x_2 \geq 10$$
$$x_1 + 2x_2 = 20$$

y

$x_2 \geq 0$ (x_1 sin restricción en el signo).

a) Use el método CER para construir el problema dual.
b) Use la tabla 6.12 para convertir el problema dual en nuestra forma estándar de la sección 6.1 y construya el problema dual correspondiente. Después muestre que este dual es equivalente al que se obtuvo en el inciso *a*).

6.4-2. Considere los problemas primal y dual en nuestra forma estándar que se presentaron en notación matricial en la sección 6.1. Use

sólo esta definición del dual de un problema primal para probar los siguientes resultados.

a) Si las restricciones funcionales del problema primal $\mathbf{Ax} \leq \mathbf{b}$ se cambian a $\mathbf{Ax} = \mathbf{b}$, el único cambio que resulta en el dual es que se eliminan las restricciones de no negatividad $\mathbf{y} \geq \mathbf{0}$. (*Sugerencia:* las restricciones $\mathbf{Ax} = \mathbf{b}$ son equivalentes al conjunto de restricciones dado por $\mathbf{Ax} \leq \mathbf{b}$ y $\mathbf{Ax} \geq \mathbf{b}$.)

b) Si las restricciones funcionales del problema primal $\mathbf{Ax} \leq \mathbf{b}$ se cambian por $\mathbf{Ax} \geq \mathbf{b}$, el único cambio que resulta en el problema dual es que las restricciones de no negatividad $\mathbf{y} \geq \mathbf{0}$ se sustituyen por restricciones de no positividad $\mathbf{y} \leq \mathbf{0}$, donde las variables duales actuales se interpretan como el negativo de las variables duales originales. (*Sugerencia*: Las restricciones $\mathbf{Ax} \geq \mathbf{b}$ son equivalentes a $-\mathbf{Ax} \leq -\mathbf{b}$.)

c) Si las restricciones de no negatividad del problema primal, $\mathbf{x} \geq \mathbf{0}$ se eliminan, el único cambio que resulta en el problema dual es que las restricciones funcionales $\mathbf{yA} \geq \mathbf{c}$ se sustituyen por $\mathbf{yA} = \mathbf{c}$. (*Sugerencia:* Una variable no restringida en signo se puede sustituir como la diferencia de dos variables no negativas.)

6.4-3.* Construya el problema dual del modelo de programación lineal que se presentó en el problema 4.6-3.

6.4-4. Considere el siguiente problema.

$$\text{Maximizar} \qquad Z = 5x_1 + 10x_2,$$

sujeta a

$$\begin{aligned} -4x_1 + 2x_2 &\geq 4 \\ 5x_1 - 10x_2 &\geq 10 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

a) Construya el problema dual.

I ***b***) Utilice el análisis gráfico del problema dual para determinar si el problema primal tiene soluciones factibles y, de ser así, si su función objetivo está acotada.

6.4-5. Considere las dos versiones del problema dual del ejemplo de terapia de radiación de las tablas 6.15 y 6.16. Repase en la sección 6.4 por qué estas dos versiones son completamente equivalentes. Complete los detalles para verificar esta equivalencia con los pasos para convertir la versión de la tabla 6.15 a formas equivalentes hasta obtener la versión de la tabla 6.16.

6.4-6. Para cada uno de los siguientes modelos de programación lineal, use el método CER para construir el dual.

a) El modelo aparece en el problema 4.6-7

b) El modelo que se presentó en el problema 4.6-16

6.4-7. Considere el modelo con restricciones de igualdad del problema 4.6-2.

a) Construya su problema dual.

b) Demuestre que la respuesta que dio en el inciso *a*) es correcta (es decir, restricciones de igualdad conducen a variables duales sin restricción de no negatividad); para ello convierta primero el problema primal a nuestra forma estándar (vea la tabla 6.12), después construya su problema dual y luego convierta este dual a la forma que obtuvo en el inciso *a*).

6.4-8.* Considere el modelo sin restricciones de no negatividad que se presenta en el problema 4.6-14.

a) Construya su problema dual.

b) Demuestre que la respuesta al inciso *a*) es correcta (es decir, variables sin restricciones de no negatividad conducen a restricciones de igualdad en el problema dual); para ello, convierta primero el problema primal a nuestra forma estándar (vea la tabla 6.12), después construya su problema dual y luego convierta éste a la forma que obtuvo en el inciso *a*).

6.4-9. Considere el problema dual del ejemplo de la Wyndor Glass Co., que se da en la tabla 6.1. Demuestre que su problema dual es el primal que se presentó en la tabla 6.1; para ello, siga los pasos de conversión de la tabla 6.13.

6.4-10. Considere el siguiente problema.

$$\text{Maximizar} \qquad Z = -5x_1 - 15x_2,$$

sujeta a

$$\begin{aligned} 2x_1 - 4x_2 &\leq 8 \\ -3x_1 + 3x_2 &\leq 24 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

I ***a***) Demuestre en una gráfica que este problema tiene una función objetivo no acotada.

b) Construya el problema dual.

I ***c***) Demuestre en forma gráfica que el problema dual no tiene soluciones factibles.

6.5-1. Considere el modelo del problema 6.7-2. Utilice la teoría de la dualidad para determinar si la solución básica actual aún es óptima después de cada uno de los siguientes cambios independientes.

a) El cambio en el inciso *e*) del problema 6.7-2

b) El cambio en el inciso *g*) del problema 6.7-2

6.5-2. Considere el modelo del problema 6.7-4. Utilice la teoría de dualidad para determinar si la solución básica actual aún es óptima después de cada uno de los siguientes cambios independientes.

a) El cambio en el inciso *b*) del problema 6.7-4

b) El cambio en el inciso *d*) del problema 6.7-4

6.5-3. Reconsidere el inciso *d*) del problema 6.7-6. Utilice la teoría de la dualidad para determinar si la solución óptima original todavía es óptima.

6.6-1.* Considere el siguiente problema.

$$\text{Maximizar} \qquad Z = 3x_1 + x_2 + 4x_3,$$

sujeta a

$$\begin{aligned} 6x_1 + 3x_2 + 5x_3 &\leq 25 \\ 3x_1 + 4x_2 + 5x_3 &\leq 20 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0.$$

el correspondiente conjunto final de ecuaciones que conduce a la solución óptima es

$$(0) \quad Z \quad + 2x_2 \quad + \frac{1}{5}x_4 + \frac{3}{5}x_5 = 17$$

$$(1) \quad x_1 - \frac{1}{3}x_2 \quad + \frac{1}{3}x_4 - \frac{1}{3}x_5 = \frac{5}{3}$$

$$(2) \quad x_2 + x_3 - \frac{1}{5}x_4 + \frac{2}{5}x_5 = 3.$$

a) Identifique la solución óptima a partir de este conjunto de ecuaciones.
b) Construya el problema dual.
I *c*) Identifique la solución óptima para el problema dual a partir del conjunto final de ecuaciones. Verifique esta solución al resolver el problema dual en forma gráfica.
d) Suponga que el problema original se cambia a

Maximizar $Z = 3x_1 + 3x_2 + 4x_3,$

sujeta a

$$6x_1 + 2x_2 + 5x_3 \leq 25$$
$$3x_1 + 3x_2 + 5x_3 \leq 20$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

Utilice la teoría de la dualidad para determinar si la solución óptima anterior todavía es óptima.
e) Use la idea fundamental que se presentó en la sección 5.3 para identificar los nuevos coeficientes de x_2 del conjunto final de ecuaciones después de que se han hecho los ajustes necesarios para los cambios que se realizaron en el problema original del inciso *d*).
f) Ahora suponga que el único cambio al problema original es la introducción de una nueva variable x_{nueva} al modelo, de la siguiente manera:

Maximizar $Z = 3x_1 + x_2 + 4x_3 + 2x_{\text{nueva}},$

sujeta a

$$6x_1 + 3x_2 + 5x_3 + 3x_{\text{nueva}} \leq 25$$
$$3x_1 + 4x_2 + 5x_3 + 2x_{\text{nueva}} \leq 20$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_{\text{nueva}} \geq 0.$$

Use la teoría de la dualidad para determinar si la solución óptima anterior, junto con $x_{\text{nueva}} = 0$, aún es óptima.
g) Utilice la idea fundamental que se presentó en la sección 5.3 para identificar los coeficientes de x_{nueva} como variable no básica del conjunto final de ecuaciones que se obtiene después de introducirla al modelo original, como se muestra en el inciso *f*).

D,I **6.6-2.** Reconsidere el modelo del problema 6.6-1. Ahora se debe llevar a cabo un análisis de sensibilidad que investigue en forma *independiente* cada uno de los siguientes cambios del modelo original. En cada cambio, utilice el procedimiento de análisis de sensibilidad para revisar el conjunto final de ecuaciones dado (en forma de tabla) y convierta este conjunto de ecuaciones a su forma apropiada de eliminación gaussiana. Después pruebe la factibilidad y la optimalidad de la solución. (No reoptimice.)

a) Cambie el lado derecho de la restricción 1 a $b_1 = 10$.
b) Cambie el lado derecho de la restricción 2 a $b_2 = 10$.
c) Cambie el coeficiente de x_2 en la función objetivo a $c_2 = 3$.
d) Cambie el coeficiente de x_3 en la función objetivo a $c_3 = 2$.
e) Cambie el coeficiente de x_2 en la restricción 2 a $a_{22} = 2$.
f) Cambie el coeficiente de x_1 en la restricción 1 a $a_{11} = 8$.

D,I **6.6-3.** Considere el siguiente problema.

Minimizar $W = 5y_1 + 4y_2,$

sujeta a

$$4y_1 + 3y_2 \geq 4$$
$$2y_1 + y_2 \geq 3$$
$$y_1 + 2y_2 \geq 1$$
$$y_1 + y_2 \geq 2$$

y

$$y_1 \geq 0, \quad y_2 \geq 0.$$

Debido a que este problema primal tiene más restricciones que variables, suponga que se aplicó el método símplex directamente a su problema dual. Si x_5 y x_6 denotan las variables de holgura de este problema dual, la tabla símplex final que resulta es

Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	x_5	x_6	Lado derecho
Z	(0)	1	3	0	2	0	1	1	9
x_2	(1)	0	1	1	−1	0	1	−1	1
x_4	(2)	0	2	0	3	1	−1	2	3

Realice un análisis de sensibilidad en cada uno de los cambios independientes en el problema primal original, investigue directamente los efectos sobre el problema dual y después infiera el efecto complementario sobre el problema primal. Aplique, a cada cambio, el procedimiento del análisis de sensibilidad que se resumió al final de la sección 6.6 al problema dual (*no* reoptimice). Después establezca las conclusiones para determinar si la solución básica actual del problema primal es aún factible y si todavía es óptima. Por último, verifique esas conclusiones con un análisis gráfico directo del problema primal.
a) Cambie la función objetivo a $W = 3y_1 + 5y_2$.
b) Cambie los lados derechos de las restricciones funcionales a 3, 5, 2 y 3, respectivamente.
c) Cambie la primera restricción a $2y_1 + 4y_2 \geq 7$.
d) Cambie la segunda restricción a $5y_1 + 2y_2 \geq 10$.

6.7-1. Lea el artículo que describe de manera completa el estudio de Investigación de Operaciones que se muestra en el Recuadro de aplicación en la sección 6.7. Describa de manera breve la forma como se aplicó el análisis de sensibilidad a este estudio. Después, haga una lista de los diferentes beneficios financieros y no financieros que resulten del estudio.

D,I **6.7-2.*** Considere el problema siguiente.

Maximizar $Z = -5x_1 + 5x_2 + 13x_3,$

sujeta a

$$-x_1 + x_2 + 3x_3 \leq 20$$
$$12x_1 + 4x_2 + 10x_3 \leq 90$$

y

$$x_j \geq 0 \qquad (j = 1, 2, 3).$$

si se establece que x_4 y x_5 son las variables de holgura de las respectivas restricciones, el método símplex conduce al siguiente conjunto final de ecuaciones:

$$\begin{array}{lrl} (0) & Z \qquad\quad + 2x_3 + 5x_4 \qquad & = 100 \\ (1) & -x_1 + x_2 + 3x_3 + x_4 \qquad & = 20 \\ (2) & 16x_1 \qquad - 2x_3 - 4x_4 + x_5 & = 10. \end{array}$$

Ahora debe llevar a cabo un análisis de sensibilidad *independiente* para investigar cada uno de los siguientes nueve cambios en el modelo original. Para cada cambio, utilice el procedimiento del análisis de sensibilidad a fin de revisar este conjunto de ecuaciones (en forma de tabla) y convertirlo en uno con la forma apropiada de eliminación de Gauss para identificar y evaluar la solución básica actual. Después pruebe la factibilidad y la optimalidad de esta solución. (No reoptimice.)

a) Cambie el lado derecho de la restricción 1 a

$$b_1 = 30.$$

b) Cambie el lado derecho de la restricción 2 a

$$b_2 = 70.$$

c) Cambie la columna del lado derecho a

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 10 \\ 100 \end{bmatrix}.$$

d) Cambie el coeficiente de x_3 en la función objetivo a

$$c_3 = 8.$$

e) Cambie los coeficientes de x_1 a

$$\begin{bmatrix} c_1 \\ a_{11} \\ a_{21} \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 5 \end{bmatrix}.$$

f) Cambie los coeficientes de x_2 a

$$\begin{bmatrix} c_2 \\ a_{12} \\ a_{22} \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \\ 5 \end{bmatrix}.$$

g) Introduzca una nueva variable x_6 con coeficientes

$$\begin{bmatrix} c_6 \\ a_{16} \\ a_{26} \end{bmatrix} = \begin{bmatrix} 10 \\ 3 \\ 5 \end{bmatrix}.$$

h) Introduzca una nueva restricción $2x_1 + 3x_2 + 5x_3 \leq 50$. (Sea x_6 su variable de holgura.)

i) Transforme la restricción 2 en

$$10x_1 + 5x_2 + 10x_3 \leq 100.$$

6.7-3.* Reconsidere el modelo del problema 6.7-2. Suponga que ahora se quiere aplicar un análisis de programación lineal paramétrica a este problema. En particular, cambie el lado derecho de las restricciones funcionales a

$$20 + 2\theta \qquad \text{(para la restricción 1)}$$

y

$$90 - \theta \qquad \text{(para la restricción 2),}$$

donde θ puede tomar cualquier valor positivo o negativo.

Exprese la solución básica (y Z) correspondiente a la solución óptima original como una función de θ. Determine las cotas superior e inferior de θ antes de que esta solución se vuelva no factible.

D,I **6.7-4.** Considere el siguiente problema.

$$\text{Maximizar} \qquad Z = 2x_1 + 7x_2 - 3x_3,$$

sujeta a

$$\begin{aligned} x_1 + 3x_2 + 4x_3 &\leq 30 \\ x_1 + 4x_2 - x_3 &\leq 10 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0.$$

Sean x_4 y x_5 las variables de holgura de las respectivas restricciones, el método símplex genera el siguiente conjunto de ecuaciones finales:

$$\begin{array}{lrl} (0) & Z + x_2 + x_3 \qquad + 2x_5 & = 20 \\ (1) & - x_2 + 5x_3 + x_4 - x_5 & = 20 \\ (2) & x_1 + 4x_2 - x_3 \qquad + x_5 & = 10. \end{array}$$

Ahora debe realizar un análisis de sensibilidad para investigar cada uno de los siguientes siete cambios *independientes* introducidos en el modelo original. En cada cambio, utilice el procedimiento de análisis de sensibilidad para revisar este conjunto de ecuaciones (en forma de tabla) y convertirlas a la forma apropiada de eliminación de Gauss a fin de identificar y evaluar la solución básica actual. Después pruebe la factibilidad y optimalidad de esta solución. Si la prueba falla, reoptimice para encontrar una nueva solución óptima.

a) Cambie los lados derechos a

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 20 \\ 30 \end{bmatrix}.$$

b) Cambie los coeficientes de x_3 a

$$\begin{bmatrix} c_3 \\ a_{13} \\ a_{23} \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \\ -2 \end{bmatrix}.$$

c) Cambie los coeficientes de x_1 a

$$\begin{bmatrix} c_1 \\ a_{11} \\ a_{21} \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}.$$

d) Introduzca una variable nueva x_6 con coeficientes

$$\begin{bmatrix} c_6 \\ a_{16} \\ a_{26} \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}.$$

e) Cambie la función objetivo a $Z = x_1 + 5x_2 - 2x_3$.

f) Introduzca una nueva restricción $3x_1 + 2x_2 + 3x_3 \leq 25$.

g) Cambie la restricción 2 a $x_1 + 2x_2 + 2x_3 \leq 35$.

6.7-5. Reconsidere el modelo del problema 6.7-4. Suponga que ahora se quiere aplicar un análisis de programación lineal paramétrica a este problema. En particular, los lados derechos de las restricciones funcionales se cambian a

$30 + 3\theta$ (para la restricción 1)

y

$10 - \theta$ (para la restricción 2),

donde θ puede tomar cualquier valor positivo o negativo.

Exprese la solución básica (y Z) correspondiente a la solución óptima original como una función de θ. Determine las cotas inferior y superior de θ antes de que esta solución se vuelva no factible.

D,I **6.7-6.** Considere el siguiente problema.

$$\text{Maximizar} \quad Z = 2x_1 - x_2 + x_3,$$

sujeta a

$$\begin{aligned} 3x_1 - 2x_2 + 2x_3 &\leq 15 \\ -x_1 + x_2 + x_3 &\leq 3 \\ x_1 - x_2 + x_3 &\leq 4 \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

Sean x_4, x_5 y x_6 las variables de holgura en las respectivas restricciones. El método símplex llevó al siguiente conjunto final de ecuaciones:

$$\begin{aligned} &(0) \quad Z + 2x_3 + x_4 + x_5 = 18 \\ &(1) \quad x_2 + 5x_3 + x_4 + 3x_5 = 24 \\ &(2) \quad 2x_3 + x_5 + x_6 = 7 \\ &(3) \quad x_1 + 4x_3 + x_4 + 2x_5 = 21. \end{aligned}$$

Ahora debe llevar a cabo un análisis de sensibilidad para investigar cada uno de los siguientes ocho cambios *independientes* introducidos en el modelo original. En cada cambio, utilice el análisis de sensibilidad para revisar este conjunto de ecuaciones (en forma de tabla) y convertirlo en uno de la forma apropiada de eliminación gaussiana para identificar y evaluar la solución básica actual. Después pruebe la factibilidad y la optimalidad de esta solución. Si cualquiera de las pruebas falla, reoptimice para encontrar una nueva solución óptima.

***a*)** Cambie el lado derecho a

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 4 \\ 2 \end{bmatrix}.$$

***b*)** Cambie el coeficiente de x_3 de la función objetivo a $c_3 = 2$.

***c*)** Cambie el coeficiente de x_1 de la función objetivo a $c_1 = 3$.

***d*)** Cambie los coeficientes de x_3 a

$$\begin{bmatrix} c_3 \\ a_{13} \\ a_{23} \\ a_{33} \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}.$$

***e*)** Cambie los coeficientes de x_1 y x_2 a

$$\begin{bmatrix} c_1 \\ a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -2 \\ 3 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} c_2 \\ a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ 3 \\ 2 \end{bmatrix},$$

respectivamente.

***f*)** Cambie la función objetivo a $Z = 5x_1 + x_2 + 3x_3$.

***g*)** Cambie la restricción 1 a $2x_1 - x_2 + 4x_3 \leq 12$.

***h*)** Introduzca una nueva restricción $2x_1 + x_2 + 3x_3 \leq 60$.

C **6.7-7.** Considere el problema de Distribution Unlimited Co., de la sección 3.13.

Aunque la figura 3.13 proporciona una estimación de los costos unitarios de envío por varias rutas, existe cierta incertidumbre sobre ellos. Por tanto, antes de adoptar la solución opcional que se presentó al final de la sección 3.4, la administración desea más información acerca del efecto de las inexactitudes en su estimación de los costos unitarios.

Use un software basado en el método símplex para generar el informe de análisis de sensibilidad en preparación para contestar las siguientes preguntas.

***a*)** ¿Cuál de los costos unitarios de envío que se proporcionan en la figura 3.13 tiene el margen de error más pequeño sin invalidar la solución óptima que se dio en la sección 3.4? ¿A qué debe dedicarse el mayor esfuerzo al estimar los costos unitarios de envío?

***b*)** ¿Cuál es el intervalo permisible para seguir óptimo de cada uno de los costos unitarios de envío?

***c*)** ¿Cómo deben interpretarse estos intervalos permisibles por la administración?

***d*)** Si estas estimaciones cambian para más de uno de los costos de envío, ¿cómo se puede usar el análisis de sensibilidad generado para determinar si la solución óptima puede cambiar?

6.7-8. Considere el siguiente problema.

$$\text{Maximizar} \quad Z = c_1x_1 + c_2x_2,$$

sujeta a

$$\begin{aligned} 2x_1 - x_2 &\leq b_1 \\ x_1 - x_2 &\leq b_2 \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

Sean x_3 y x_4 las variables de holgura de las restricciones respectivas. Cuando $c_1 = 3$, $c_2 = -2$, $b_1 = 30$ y $b_2 = 10$, el método símplex llevó a la siguiente tabla símplex final.

Variable básica	Ec.	Coeficiente de:					Lado derecho
		Z	x_1	x_2	x_3	x_4	
Z	(0)	1	0	0	1	1	40
x_2	(1)	0	0	1	1	−2	10
x_1	(2)	0	1	0	1	−1	20

I ***a***) Use el análisis gráfico para determinar el intervalo permisible para que la solución de la tabla siga siendo óptima con c_1 y c_2.

b) Use el análisis algebraico para derivar y verificar sus respuestas al inciso *a*).

I ***c***) Use el análisis gráfico para determinar el intervalo permisible para seguir factible de b_1 y b_2.

d) Use el análisis algebraico para derivar y verificar sus respuestas del inciso *c*)

C ***e***) Use un software basado en el método símplex para encontrar los intervalos permisibles.

I **6.7-9.** Considere la variación 5 del modelo de Wyndor Glass Co., (vea la figura 6.6 y la tabla 6.24), donde los cambios en los valores de los parámetros dados en la tabla 6.21 son $\bar{c}_2 = 3$, $\bar{a}_{22} = 3$ y $\bar{a}_{32} = 4$. Use la fórmula $\mathbf{b}^* = \mathbf{S}^* \overline{\mathbf{b}}$ para encontrar el rango permisible de cada b_i. Después interprete cada rango permisible en forma gráfica.

I **6.7-10.** Considere la variación 5 del modelo de Wyndor Glass Co. (vea la figura 6.6 y la tabla 6.24), donde los cambios en los valores de los parámetros dados en la tabla 6.21 son $\bar{c}_2 = 3$, $\bar{a}_{22} = 3$, y $\bar{a}_{32} = 4$. Verifique algebraica y gráficamente que el intervalo permisible para para que c_1 continúe siendo óptimo es $c_1 \geq \frac{9}{4}$.

6.7-11. En el problema de la tabla 6.21 encuentre el intervalo permisible de c_2 para seguir óptimo. Muestre su trabajo algebraico y use para ello la tabla símplex de la tabla 6.21. Después justifique su respuesta desde el punto de vista geométrico, con referencia a la figura 6.3.

6.7-12.* En el problema original de la Wyndor Glass Co., utilice la última tabla símplex de la tabla 4.8 para hacer lo siguiente.

a) Encuentre el intervalo permisible para que cada b_i continúe siendo factible.

b) Encuentre el intervalo permisible de c_1 y c_2 para seguir óptimos.

C ***c***) Use un paquete de software basado en el método símplex para encontrar los intervalos permisibles.

6.7-13. En el caso de la variación 6 del modelo de la Wyndor Glass Co., que se presentó en la sección 6.7, utilice la última tabla símplex de la tabla 6.25 para hacer lo siguiente.

a) Encuentre el intervalo permisible para que cada b_i siga factible.

b) Encuentre el intervalo permisible para que c_1 y c_2 sigan óptimos.

C ***c***) Use un paquete de software basado en el método símplex para encontrar los intervalos permisibles.

6.7-14. Considere la variación 5 del modelo de la Wyndor Glass Co., que se presentó en la sección 6.7, donde $\bar{c}_2 = 3$, $\bar{a}_{22} = 3$ y $\bar{a}_{32} = 4$, y donde los otros parámetros se presentan en la tabla 6.21. A partir de la tabla símplex final de la parte inferior de la tabla 6.24, construya una tabla como la 6.26 para realizar un análisis de programación lineal paramétrica, donde

$$c_1 = 3 + \theta \qquad \text{y} \qquad c_2 = 3 + 2\theta.$$

¿Cuánto puede aumentar el valor de θ antes de que la solución básica actual deje de ser óptima?

6.7-15. Reconsidere el modelo del problema 6.7-6. Suponga que ahora existe la opción de hacer trueques en las ganancias que se obtienen de las dos primeras actividades, es decir, el coeficiente de x_1 de la función objetivo se puede aumentar en cualquier cantidad si a la vez se disminuye el de x_2 en la misma cantidad. Así, las alternativas que se tienen para la función objetivo son

$$Z(\theta) = (2 + \theta)x_1 - (1 + \theta)x_2 + x_3,$$

donde puede tomar cualquier valor no negativo.

Construya una tabla como la 6.26 para realizar un análisis de programación lineal paramétrica de este problema. Determine la cota superior de θ antes de que la solución óptima original deje de ser óptima. Después determine el mejor valor de θ en este intervalo.

6.7-16. Considere el siguiente problema de programación lineal paramétrica.

$$\text{Maximizar} \qquad Z(\theta) = (10 - 4\theta)x_1 + (4 - \theta)x_2 + (7 + \theta)x_3,$$

sujeta a

$$3x_1 + x_2 + 2x_3 \leq 7 \qquad \text{(recurso 1)},$$
$$2x_1 + x_2 + 3x_3 \leq 5 \qquad \text{(recurso 2)},$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0,$$

donde θ puede tomar cualquier valor positivo o negativo. Sean x_4 y x_5 las variables de holgura de las restricciones respectivas. Después de aplicar el método símplex con $\theta = 0$, la tabla símplex final es

Variable básica	Ec.	Coeficiente de:						Lado derecho
		Z	x_1	x_2	x_3	x_4	x_5	
Z	(0)	1	0	0	3	2	2	24
x_1	(1)	0	1	0	−1	1	−1	2
x_2	(2)	0	0	1	5	−2	3	1

a) Determine el intervalo de valores de θ para el que la solución BF anterior permanece óptima. Después encuentre la mejor elección de θ dentro de este intervalo.

b) Dado que θ está dentro del intervalo que se encontró en el inciso *a*), encuentre el intervalo permisible de b_1 (la cantidad disponible del recurso 1) para que continúe siendo factible. Después haga lo mismo en relación con b_2 (la cantidad disponible del recurso 2).

c) Dado que θ se encuentra dentro del intervalo que se encontró en el inciso *a*), identifique los precios sombra (como función de θ) de los dos recursos. Utilice esta información para determinar cómo cambia la función objetivo (como función de θ) si disminuye la cantidad disponible del recurso 1 en una unidad y al mismo tiempo aumenta la cantidad disponible del recurso 2 en una unidad.

d) Construya el dual de este problema de programación lineal paramétrica. Haga $\theta = 0$ y resuelva en forma gráfica este problema dual para encontrar los precios sombra correspondientes a los dos recursos del problema primal. Después encuentre estos precios sombra como función de θ [dentro del intervalo de valores que se encontró en el inciso *a*)] con la obtención algebraica de esta misma solución FEV óptima para el problema dual como función de θ.

6.7-17. Considere el siguiente problema de programación lineal paramétrica.

$$\text{Maximizar} \qquad Z(\theta) = 2x_1 + 4x_2 + 5x_3,$$

sujeta a

$$x_1 + 3x_2 + 2x_3 \le 5 + \theta$$
$$x_1 + 2x_2 + 3x_3 \le 6 + 2\theta$$

y

$$x_1 \ge 0, \quad x_2 \ge 0, \quad x_3 \ge 0,$$

donde θ puede tomar cualquier valor positivo o negativo. Sean x_4 y x_5 las variables de holgura de las restricciones funcionales respectivas. Después de aplicar el método símplex con $\theta = 0$, la tabla símplex final es

Variable básica	Ec.	Coeficiente de:						Lado derecho
		Z	x_1	x_2	x_3	x_4	x_5	
Z	(0)	0	0	1	0	1	1	11
x_1	(1)	1	1	5	0	3	−2	3
x_3	(2)	2	0	−1	1	−1	1	1

a) Exprese la solución BF (y Z) que se presenta en esta tabla, como una función de θ. Determine las cotas superior e inferior de θ antes de que esta solución óptima se vuelva no factible. Después determine la mejor elección de θ entre estas cotas.

b) Dado que θ está dentro de las cotas que se encontraron en el inciso *a*), determine el intervalo permisible para que c_1 (coeficiente de x_1 en la función objetivo) continúe siendo óptimo.

6.7-18. Considere el siguiente problema

$$\text{Maximizar} \quad Z = 10x_1 + 4x_2,$$

sujeta a

$$3x_1 + x_2 \le 30$$
$$2x_1 + x_2 \le 25$$

y

$$x_1 \ge 0, \quad x_2 \ge 0.$$

Sean x_3 y x_4 las variables de holgura de las respectivas restricciones funcionales. Después de aplicar el método símplex, la tabla final es

Variable básica	Ec.	Coeficiente de:					Lado derecho
		Z	x_1	x_2	x_3	x_4	
Z	(0)	1	0	0	2	2	110
x_2	(1)	0	0	1	−2	3	15
x_1	(2)	0	1	0	1	−1	5

Ahora suponga que se introducen, al mismo tiempo, los dos cambios siguientes en el modelo original:

1. La primera restricción se cambia a $4x_1 + x_2 \le 40$.
2. Se introduce la programación lineal paramétrica para cambiar la función objetivo a las opciones alternativas de

$$Z(\theta) = (10 - 2\theta)x_1 + (4 + \theta)x_2,$$

donde puede elegirse cualquier valor no negativo de θ.

a) Construya la tabla símplex final revisada que resulte (como función de θ), y después convierta esta tabla símplex a la forma apropiada de eliminación de Gauss. Utilice esta tabla para identificar la nueva solución óptima que se obtiene ya sea para $\theta = 0$ o para valores suficientemente pequeños de θ.

b) ¿Cuál es la cota superior sobre θ antes de que esta solución deje de ser óptima?

c) Para el intervalo de θ de cero a su cota superior, ¿cuál es el valor de θ que da el mayor valor de la función objetivo?

6.7-19. Considere el siguiente problema.

$$\text{Maximizar} \quad Z = 9x_1 + 8x_2 + 5x_3,$$

sujeta a

$$2x_1 + 3x_2 + x_3 \le 4$$
$$5x_1 + 4x_2 + 3x_3 \le 11$$

y

$$x_1 \ge 0, \quad x_2 \ge 0, \quad x_3 \ge 0.$$

Sean x_4 y x_5 las variables de holgura de las restricciones funcionales respectivas. Después de aplicar el método símplex, la tabla símplex final es

Variable básica	Ec.	Coeficiente de:						Lado derecho
		Z	x_1	x_2	x_3	x_4	x_5	
Z	(0)	1	0	2	0	2	1	19
x_1	(1)	0	1	5	0	3	−1	1
x_3	(2)	0	0	−7	1	−5	2	2

D,I ***a***) Suponga que ahora se dispone de una nueva tecnología para llevar a cabo la primera actividad que se considera en este problema. Si se adoptara esta nueva tecnología en lugar de la que se tiene, los coeficientes de x_1 en el modelo cambiarían

$$\text{de} \quad \begin{bmatrix} c_1 \\ a_{11} \\ a_{21} \end{bmatrix} = \begin{bmatrix} 9 \\ 2 \\ 5 \end{bmatrix} \quad \text{a} \quad \begin{bmatrix} c_1 \\ a_{11} \\ a_{21} \end{bmatrix} = \begin{bmatrix} 18 \\ 3 \\ 6 \end{bmatrix}.$$

Utilice el procedimiento de análisis de sensibilidad para investigar el efecto potencial y cuán deseable sería adoptar la nueva tecnología. En particular, suponga que sí se adoptará y construya la tabla símplex final; luego, convierta esta tabla a la forma apropiada de eliminación de Gauss y después reoptimice (si es necesario) para encontrar la nueva solución óptima.

b) Ahora suponga que se tiene la opción de mezclar la tecnología anterior y la nueva para realizar la primera actividad. Sea θ la porción de la nueva tecnología que se incluye, de manera que $0 \le \theta \le 1$. Dado θ, los coeficientes de x_1 en el modelo cambian a

$$\begin{bmatrix} c_1 \\ a_{11} \\ a_{21} \end{bmatrix} = \begin{bmatrix} 9 + 9\theta \\ 2 + \theta \\ 5 + \theta \end{bmatrix}.$$

Construya la tabla símplex final revisada (como una función de θ) y conviértala a la forma apropiada de eliminación de Gauss. Utilice esta tabla símplex para identificar la solución básica actual como una función de θ. Para los valores permitidos de $0 \le \theta \le 1$, proporcione

un intervalo de θ valores para los que esta solución es tanto factible como óptima. ¿Cuál es la mejor elección para el valor de θ dentro de este intervalo?

6.7-20. Considere el siguiente problema.

Maximizar $Z = 3x_1 + 5x_2 + 2x_3,$

sujeta a

$$-2x_1 + 2x_2 + x_3 \le 5$$
$$3x_1 + x_2 - x_3 \le 10$$

y

$$x_1 \ge 0, \quad x_2 \ge 0, \quad x_3 \ge 0.$$

Sean x_4 y x_5 las variables de holgura de las restricciones funcionales respectivas. Después de aplicar el método símplex, la tabla símplex final es

Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	x_5	Lado derecho
Z	(0)	1	0	20	0	9	7	115
x_1	(1)	0	1	3	0	1	1	15
x_3	(2)	0	0	8	1	3	2	35

Ahora debe aplicarse programación lineal paramétrica en forma simultánea a la función objetivo y a los lados derechos, donde el modelo en términos del nuevo parámetro es el siguiente:

Maximizar $Z(\theta) = (3 + 2\theta)x_1 + (5 + \theta)x_2 + (2 - \theta)x_3,$

sujeta a

$$-2x_1 + 2x_2 + x_3 \le 5 + 6\theta$$
$$3x_1 + x_2 - x_3 \le 10 - 8\theta$$

y

$$x_1 \ge 0, \quad x_2 \ge 0, \quad x_3 \ge 0.$$

Construya la tabla símplex final revisada (como función de θ) y conviértala a la forma apropiada de eliminación de Gauss. Utilice esta tabla símplex para identificar la solución básica actual como función de θ. Para $\theta \ge 0$, dé el intervalo de valores de θ donde esta solución es factible y óptima. ¿Cuál es el mejor valor de θ en este intervalo?

6.7-21. Considere el problema de la Wyndor Glass Co., descrito en la sección 3.1. Suponga que, además del estudio de la introducción de dos nuevos productos, la administración también pretende cambiar la tasa de producción de un antiguo producto que todavía es redituable. Consulte la tabla 3.1. La capacidad utilizada por unidad de tasa de producción de este antiguo producto es 1, 4 y 3 en las plantas 1, 2 y 3, respectivamente. Por tanto, si θ denota el *cambio* (positivo o negativo) de esta tasa de producción, el lado derecho de las tres restricciones funcionales en la sección 3.1 se convierte en $4 - \theta$, $12 - 4\theta$ y $18 - 3\theta$, respectivamente. Así, un valor negativo de θ libera capacidad adicional para producir más de los dos nuevos productos, mientras que un valor positivo tiene el efecto opuesto.

***a*)** Utilice la formulación de programación lineal paramétrica para determinar el efecto de los distintos valores de θ sobre la solución óptima de la mezcla de los dos nuevos productos que se presentó en la tabla símplex final de la tabla 4.8. En particular, utilice la idea fundamental de la sección 5.3 para obtener expresiones de Z y de las variables básicas x_3, x_2 y x_1 en términos de θ, y suponga que θ es lo suficientemente cercano a cero como para que esta solución básica "final" todavía sea factible y, por tanto, óptima para el valor dado de θ.

***b*)** Ahora considere una pregunta más amplia respecto de la elección de la mezcla de los dos nuevos productos. ¿Cuál es la ganancia unitaria de punto de equilibrio del producto antiguo (en comparación con los dos nuevos productos), bajo la cual debe disminuirse su tasa de producción ($\theta < 0$) en favor de los dos nuevos productos y por encima de la cual debe aumentarse ésta ($\theta > 0$)?

***c*)** Si la ganancia unitaria se encuentra por encima de este punto de equilibrio, ¿cuánto puede disminuir la tasa de producción de este producto antiguo antes de que la solución BF final pierda factibilidad?

***d*)** Si la ganancia unitaria es menor que este punto de equilibrio, ¿cuánto puede disminuir su tasa de producción (si se supone que la tasa anterior era más alta que esta disminución) antes de que la solución BF final pierda factibilidad?

6.7-22. Considere el siguiente problema.

Maximizar $Z = 2x_1 - x_2 + 3x_3,$

sujeta a

$$x_1 + x_2 + x_3 = 3$$
$$x_1 - 2x_2 + x_3 \ge 1$$
$$2x_2 + x_3 \le 2$$

y

$$x_1 \ge 0, \quad x_2 \ge 0, \quad x_3 \ge 0.$$

Suponga que se empleó el método de la gran M (sección 4.6) para obtener la solución BF inicial (artificial). Sea la variable $\bar{x}_4$ de holgura artificial de la primera restricción, x_5 la variable de superávit de la segunda restricción, $\bar{x}_6$ la variable artificial de la segunda restricción y x_7 la variable de holgura de la tercera restricción. El conjunto de ecuaciones final correspondiente que proporciona la solución óptima es

$$(0) \quad Z + 5x_2 + (M+2)\bar{x}_4 + M\bar{x}_6 + x_7 = 8$$
$$(1) \quad x_1 - x_2 + \bar{x}_4 - x_7 = 1$$
$$(2) \quad 2x_2 + x_3 + x_7 = 2$$
$$(3) \quad 3x_2 + \bar{x}_4 + x_5 - \bar{x}_6 = 2.$$

Suponga que la función objetivo original se cambia a $Z = 2x_1 + 3x_2 + 4x_3$ y que la tercera restricción original se cambia a $2x_2 + x_3 \le 1$. Utilice el procedimiento de análisis de sensibilidad para revisar el conjunto de ecuaciones final (en forma de tabla) y convertirlo a la forma apropiada de eliminación de Gauss para identificar y evaluar la solución básica actual. Después pruebe la factibilidad y la optimalidad de esta solución. (No reoptimice.)

6.8-1. Considere el siguiente problema.

Maximizar $Z = 2x_1 + 5x_2,$

sujeta a

$$x_1 + 2x_2 \leq 10 \text{ (recurso 1)}$$
$$x_1 + 3x_2 \leq 12 \text{ (recurso 2)}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0,$$

donde Z mide la ganancia en dólares proveniente de las dos actividades.

Mientras se hacía un análisis de sensibilidad, se supo que las estimaciones de las ganancias unitarias tienen una exactitud de sólo ±50%. En otras palabras, los rangos de los *valores probables* de estas ganancias unitarias son de \$1 a \$3 en el caso de la actividad 1 y de \$2.50 a \$7.50 de la actividad 2.

E* ***a***) Formule un modelo en hoja de cálculo para este problema basado en las estimaciones iniciales de las ganancias unitarias. Después, use el Solver para encontrar una solución óptima y generar el informe de sensibilidad.

E* ***b***) Use la hoja de cálculo y el Solver para verificar si esta solución óptima aún lo es con cambios en la ganancia unitaria de la actividad 1 de \$2 a \$1 y de \$2 a \$3.

E* ***c***) También verifique si la solución óptima mantiene su carácter de tal si la ganancia unitaria de la actividad 1 permanece en \$2 pero para la actividad 2 cambia de \$5 a \$2.50, y de \$5 a \$7.50.

E* ***d***) Use la tabla de Solver para generar de manera sistemática la solución óptima y la ganancia total cuando la ganancia unitaria de la actividad 1 se incrementa de \$1 a \$3 en intervalos de 20 centavos (sin cambiar la ganancia unitaria de la actividad 2). Después haga lo mismo cuando la ganancia unitaria de la actividad 2 se incrementa de \$5 a \$7.50 en intervalos de 50 centavos (sin que cambie la ganancia unitaria de la actividad 1). Use estos resultados para estimar el rango permisible para que la ganancia unitaria de cada actividad continúe siendo óptima.

I ***e***) Use la rutina del método gráfico y análisis de sensibilidad del IOR Tutorial para estimar el rango permisible para que la ganancia unitaria de cada actividad conserve su optimalidad.

E* ***f***) Use el informe de sensibilidad que proporciona el Solver de Excel para encontrar el rango permisible para que la ganancia unitaria de cada actividad conserve su optimalidad. Después, use estos intervalos para verificar sus resultados en los incisos *b*) al *e*).

E* ***g***) Use la Tabla de Solver de dos vías para generar de manera sistemática la solución óptima cuando las ganancias unitarias de las dos actividades cambian al mismo tiempo, tal como se describe en el inciso *d*).

I ***h***) Use la rutina del método gráfico y análisis de sensibilidad del IOR Tutorial para interpretar en forma gráfica los resultados del inciso *g*).

E* **6.8-2.** Reconsidere el modelo que se presentó en el problema 6.8-1. Mientras se hacía un análisis de sensibilidad, se supo que las estimaciones de los lados derechos de las dos restricciones funcionales tienen una exactitud de sólo ±50%. En otras palabras, los rangos de los *valores probables* para estos parámetros son de 5 a 15 para el primer lado derecho y de 6 a 18 para el segundo.

a) Después de resolver el modelo en hoja de cálculo original, determine el precio sombra de la primera restricción funcional al incrementar su lado derecho en 1 y resuelva de nuevo.

b) Utilice la Tabla de Solver para generar la solución óptima y la ganancia total cuando el lado derecho de la primera restricción funcional se incrementa de 5 a 15 en intervalos de 1. Use esta tabla para estimar el rango permisible para este lado derecho, es decir, el rango en el que el precio sombra que obtuvo en el inciso *a*) es válido.

c) Repita el inciso *a*) para la segunda restricción funcional.

d) Repita el inciso *b*) para la segunda restricción funcional donde su lado derecho se incrementa de 6 a 18 en intervalos de 1.

e) Utilice el informe de sensibilidad del Solver para determinar el precio sombra de cada restricción funcional y el rango permisible para el lado derecho de cada una de estas restricciones.

6.8-3. Considere el siguiente problema.

$$\text{Maximizar} \qquad Z = x_1 + 2x_2,$$

sujeta a

$$x_1 + 3x_2 \leq 8 \text{ (recurso 1)}$$
$$x_1 + x_2 \leq 4 \text{ (recurso 2)}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0,$$

donde Z mide la ganancia en dólares proveniente de las dos actividades y los lados derechos se refieren al número de unidades disponibles de los recursos respectivos.

I ***a***) Aplique el método gráfico para resolver este problema.

I ***b***) Utilice el análisis gráfico para determinar el precio sombra de cada uno de estos recursos al resolver de nuevo después de incrementar la cantidad del recurso disponible en 1.

E* ***c***) Use la hoja de cálculo y el Solver para resolver los incisos *a*) y *b*).

E* ***d***) En el caso de cada uno de los recursos y de forma independiente, utilice la Tabla de Solver para generar de manera sistemática la solución óptima y la ganancia total cuando el único cambio es que la cantidad de ese recurso disponible crece en intervalos de 1, desde 4 por debajo del valor original hasta 6 por encima de ese valor. Use estos resultados para estimar el rango permisible para la cantidad disponible de cada recurso.

e) Use el informe de sensibilidad del Solver para obtener los precios sombra, así como para encontrar el rango de la cantidad disponible de cada recurso dentro del cual el precio sombra correspondiente permanece válido.

f) Describa por qué estos precios sombra son útiles cuando la administración tiene flexibilidad suficiente para cambiar las cantidades de los recursos que estarán disponibles.

6.8-4.* Uno de los productos de la G. A. Tanner Company es un tipo especial de juguete que proporciona un ganancia unitaria estimada de \$3. Debido a la gran demanda, la administración desea aumentar su tasa de producción del nivel actual de 1 000 por día. Sin embargo, un abastecimiento limitado de dos subensambles (A y B) lo dificulta. Cada juguete requiere de dos subensambles tipo A, pero el proveedor sólo puede aumentar la tasa de entrega de los 2 000 por día actuales a un máximo de 3 000. Cada juguete requiere de sólo un subensamble

tipo B, y el proveedor no podrá aumentar la tasa de entrega más allá del nivel actual de 1 000 por día. Como no dispone de otros proveedores, la administración piensa iniciar un nuevo proceso de producción interno para fabricar al mismo tiempo igual número de subensambles de los dos tipos para completar las entregas de los dos proveedores. Se estima que el costo de producir un subensamble de cada tipo será de $2.50 más que el costo de compra. La administración desea determinar la tasa de producción del juguete y la de cada par de subensambles (uno A y uno B) que maximizará la ganancia total.

La siguiente tabla resume los datos del problema.

	Utilización de recurso por unidad de cada actividad		
	Actividad		
Recurso	**Producir juguetes**	**Producir subensambles**	**Cantidad de recursos disponibles**
Subensamble A	2	–1	3 000
Subensamble B	1	–1	1 000
Ganan. unitaria	$3	–$2.50	

E* ***a***) Formule y resuelva un modelo en hoja de cálculo de este problema.

E* ***b***) Como las ganancias unitarias establecidas son sólo estimaciones, la administración desea saber cuánto puede estar equivocada cada una antes de que cambie la solución óptima. Comience por examinar esta cuestión con respecto a la primera actividad (producción de juguetes) con el uso de una hoja de cálculo y el Solver para generar de forma manual una tabla que proporcione la solución óptima y la ganancia total si la ganancia unitaria de esta actividad aumenta en incrementos de ¢50, de $2.00 a $4.00. ¿Cuánto puede diferir de su valor original de $3.00 la estimación de esta ganancia unitaria en cada dirección antes de que cambie la solución óptima?

E* ***c***) Repita el inciso *b*) con relación a la segunda actividad (producir subensambles), y resuelva de nuevo con incrementos de ¢50 de la ganancia unitaria, de $-\$3.50$ a $-\$1.50$ (con la ganancia unitaria fija para la primera actividad de $3).

E* ***d***) Utilice la Tabla de Solver para generar sistemáticamente todos los datos que se piden en los incisos *b*) y *c*), sólo que ahora utilice incrementos de ¢25 en lugar de ¢50. Use estos datos para perfeccionar sus conclusiones de los incisos *b*) y *c*).

I ***e***) Emplee la rutina del método gráfico y el análisis de sensibilidad del IOR Tutorial para determinar cuánto puede cambiar la ganancia unitaria de cada actividad en cualquier dirección (sin que cambie la ganancia unitaria de la otra actividad) antes de que la solución óptima cambie. Use esta información para especificar el rango permisible para que la ganancia unitaria de cada actividad conserve su optimalidad.

E* ***f***) Use el informe de sensibilidad de Excel y encuentre el rango permisible para que la ganancia unitaria de cada actividad conserve su optimalidad.

E* ***g***) Use la Tabla de Solver de dos vías para generar de manera sistemática la solución óptima cuando las ganancias unitarias de las dos actividades se cambian al mismo tiempo como se describe en los incisos *b*) y *c*).

h) Use la información que proporciona el informe de sensibilidad de Excel para describir cuánto pueden cambiar las ganancias unitarias de las dos actividades de manera simultánea antes de que la solución óptima cambie.

E* **6.8-5.** Reconsidere el problema 6.8-4. Después de otras negociaciones con cada proveedor, la administración de G. A. Tanner Co., descubrió que estarían dispuestos a considerar un aumento en la entrega de sus respectivos subensambles sobre los máximos establecidos (3 000 del tipo A y 1 000 del B por día) a cambio de un pequeño aumento del precio, respecto del precio normal, en los subensambles adicionales. Todavía se puede negociar el aumento de precio. La demanda del juguete es suficiente para vender 2 500 unidades por día si aumenta la entrega de subensambles para cumplir con esta tasa de producción. Suponga que las estimaciones originales de las ganancias unitarias que se presentaron en el problema 6.8-4 son exactas.

a) Formule y resuelva un modelo en hoja de cálculo para este problema con los niveles de entrega máximos originales y la restricción adicional de no producir más de 2 500 juguetes por día.

b) Sin considerar el costo adicional, utilice la hoja de cálculo y el Solver para determinar los precios sombra de la restricción del subensamble A cuando resuelva de nuevo el modelo después de aumentar la entrega máxima en 1. Básese en este precio sombra para determinar el costo adicional máximo que la compañía debe pagar por cada subensamble de este tipo.

c) Repita el inciso *b*) para la restricción del subensamble B.

d) Estime cuánto puede aumentar el envío máximo de subensambles tipo A antes de que los precios sombra (y el costo adicional) que se determinaron en el inciso *b*) ya no sean válidos; utilice la Tabla de Solver para generar una solución óptima y la ganancia total (sin costo adicional) cuando el envío aumenta en incrementos de 100, desde 3 000 hasta 4 000.

e) Repita el inciso *d*) para los subensambles tipo B, cuando el envío aumente en incrementos de 100, desde 1 000 hasta 2 000.

f) Use el informe de sensibilidad del Solver para determinar el precio sombra de cada una de las restricciones de subensamble y el rango permisible para el lado derecho de cada una de estas restricciones.

E* **6.8-6.*** Considere el problema de Union Airways de la sección 3.4, incluso los datos que se dieron en la tabla 3.19. Los archivos de Excel del capítulo 3 incluyen una hoja de cálculo que muestra la formulación y la solución óptima de este problema. Utilice dicha hoja de cálculo y el Solver de Excel para resolver los incisos *a*) al *g*) que se enuncian a continuación.

La administración está por iniciar las negociaciones de un nuevo contrato con el sindicato de agentes de servicio a clientes de la compañía. Esto puede causar cambios pequeños en los costos diarios por agente que se presentaron en la tabla 3.19 de los diferentes turnos. Se consideran por separado varios cambios posibles enumerados a continuación. En cada caso, la administración querrá saber si causará que la solución (dada en la sección 3.4) no seguirá óptima. Conteste esta pregunta en los incisos *a*) al *e*) directamente mediante la hoja de cálculo y el Solver. Si la solución óptima cambia, registre la nueva solución.

a) El costo diario por agente del turno 2 cambia de $160 a $165.

b) El costo diario por agente del turno 4 cambia de $180 a $170.

c) Ocurren tanto los cambios en *a*) como en *b*).

d) El costo diario por agente aumenta $4 en los turnos 2, 4 y 5, pero disminuye $4 en los turnos 1 y 3.

e) El costo diario por agente aumenta 2% en cada turno.

f) Use el Solver para generar el informe de análisis de sensibilidad de este problema. Suponga que los cambios anteriores se consideran después sin tener el modelo en hoja de cálculo disponible de inmediato en una computadora. Muestre en cada caso cómo se puede usar el informe de sensibilidad para verificar si la solución óptima original aún conserva su optimalidad.

g) Con respecto a cada uno de los cinco turnos independientes, use la Tabla de Solver para generar en forma sistemática la solución óptima y el costo total cuando la única modificación adicional es que el costo diario por agente de ese turno aumenta en incrementos de $3 desde $15 por debajo del costo actual hasta $15 por encima de él.

E* **6.8-7.** Reconsidere el problema de Union Airways y su modelo en hoja de cálculo que se utilizó en el problema 6.8-6.

Ahora la administración piensa aumentar el nivel de servicio a los clientes con el incremento de uno o más de los números de la columna de la derecha del número mínimo de agentes necesarios en los diferentes periodos. Como guía para tomar esta decisión, desean saber qué efecto tendrá este cambio en el costo total.

Use el Excel Solver para generar el análisis de sensibilidad como preparación para contestar las siguientes preguntas.

a) ¿Cuál de los números de la columna de la derecha de la tabla 3.19 puede aumentar sin que aumente el costo total? En cada caso indique cuál es el aumento (si es el único cambio) antes de que aumente el costo total.

b) En el caso de cada uno de los otros números, ¿cuánto aumentaría el costo total debido a un incremento de 1 en el número? En cada respuesta indique cuánto puede aumentar el número (si es el único cambio) antes de que la respuesta no sea válida.

c) ¿Siguen válidas las respuestas del inciso *b*) si todos los números considerados en ese inciso aumentan 1 al mismo tiempo?

d) ¿Cuánto pueden aumentar los 10 números de manera simultánea antes de que las respuestas del inciso *b*) ya no sean válidas?

e) ¿Siguen válidas las respuestas del inciso *b*) si los 10 números aumentan 1 al mismo tiempo?

6.8-8. David, Diana y Lidia son los únicos socios y empleados de una compañía que produce relojes finos. David y Diana pueden trabajar un máximo de 40 horas por semana, mientras que Lidia sólo puede trabajar hasta 20 horas semanales.

La empresa hace dos tipos de relojes: el reloj de pedestal y el de pared. Para hacer un reloj, David (ingeniero mecánico) ensambla las partes internas y Diana (ebanista) produce las cajas de madera labradas a mano. Lidia es responsable de recibir pedidos y enviar los relojes. El tiempo que se requiere para cada tarea se muestra en la tabla.

	Tiempo requerido	
Tarea	**Reloj de pedestal**	**Reloj de pared**
Ensamblar mecanismo del reloj	6 horas	4 horas
Tallar la cubierta de madera	8 horas	4 horas
Envío	3 horas	3 horas

Cada reloj de pedestal construido y enviado deja una ganancia de $300, mientras que cada reloj de pared proporciona una ganancia de $200.

Los tres socios desean determinar cuántos relojes de cada tipo deben producir por semana para maximizar la ganancia total.

a) Formule un modelo de programación lineal en forma algebraica para resolver este problema.

I *b*) Utilice la rutina de método gráfico y análisis de sensibilidad del IOR Tutorial para resolver el modelo. Después use esta rutina para determinar si la solución óptima cambia cuando la estimación de la ganancia unitaria por reloj de pedestal se incrementa de $300 a $375 (sin otro cambio en el modelo). Después verifique si la solución óptima se modificaría si, además de este cambio en la ganancia unitaria del reloj de pedestal, la ganancia unitaria estimada de los relojes de pared también cambia de $200 a $175.

E* *c*) Formule y resuelva este modelo en una hoja de cálculo.

E* *d*) Use el Excel Solver para verificar el efecto de los cambios especificados en el inciso *b*).

E* *e*) Use la tabla de Solver para generar de manera sistemática la solución óptima y la ganancia total cuando la ganancia unitaria de los relojes de pedestal aumenta en incrementos de $20 desde $150 hasta $450 (sin cambio en la ganancia unitaria de los relojes de pared). Después haga lo mismo cuando la ganancia unitaria de los relojes de pared aumente en incrementos de $20 desde $50 hasta $350 (sin cambio en la ganancia unitaria de los relojes de pedestal). Utilice esta información para estimar el rango permisible para que las ganancias unitarias de cada tipo de reloj mantengan su carácter de óptimas.

E* *f*) Use la Tabla de Solver de dos vías para generar en forma sistemática la solución óptima cuando las ganancias unitarias de los dos tipos de relojes se cambian simultáneamente como se especifica en el inciso *e*); utilice incrementos de $50 en lugar de $20.

E* *g*) Para el caso de los tres socios independientes, use el Solver de Excel para determinar el efecto sobre la solución óptima y la ganancia total si sólo cada uno de ellos aumenta en 5 el número máximo de horas disponibles por semana para trabajar.

E* *h*) Utilice la tabla de Solver para generar en forma sistemática la solución óptima y la ganancia total cuando el único cambio es que el número máximo de horas disponibles por semana para trabajar de David cambia a cada uno de los siguientes valores: 35, 37, 39, 41, 43, 45. Después haga lo mismo cuando el único cambio es que los números de Diana cambian de la misma forma. Por último, repita el ejercicio cuando el único cambio es que el número máximo de horas disponibles por semana para trabajar de Lidia cambia a cada uno de los siguientes valores: 15, 17, 19, 21, 23, 25.

E* *i*) Genere el informe de sensibilidad de Excel y úselo para determinar el rango permisible para que la ganancia unitaria de cada tipo de reloj y el rango permisible del máximo número de horas disponibles que cada socio tiene para trabajar por semana conserve su optimalidad.

j) Para aumentar la ganancia total, los tres socios acordaron que uno de ellos aumentaría un poco el número máximo de horas de trabajo por semana. La elección se basa en quién aumentaría más la ganancia total. Use el informe de sensibilidad para realizar la elección. (Suponga que no hay cambio en las estimaciones originales de las ganancias unitarias.)

***k*)** Explique por qué un precio sombra es igual a cero.

***l*)** ¿Es válido usar los precios sombra que se proporcionaron en el informe de sensibilidad para determinar el efecto si Lidia cambiara su máximo de horas semanales de 20 a 25? Si es así, ¿cuál sería el incremento de la ganancia total?

***m*)** Repita el inciso *l*) si, además del cambio de Lidia, también David cambiara su máximo de horas semanales de 40 a 35.

I ***n*)** Utilice el análisis gráfico para verificar la respuesta de *m*).

CASOS

CASO 6.1 Control de la contaminación

Consulte el problema de la Nori & Leets Co., en la sección 3.4 (subsección llamada "Control de la contaminación"). Allí se mencionó que una vez que el equipo de IO obtuvo una solución óptima, llevó a cabo un análisis de sensibilidad. Se solicita al lector que reconstruya los pasos que siguió el equipo de IO, después de proporcionarle algunos antecedentes adicionales.

Los valores de los parámetros para formular el modelo original se presentaron en las tablas 3.12, 3.13 y 3.14. Como la compañía no tiene experiencia previa en métodos de control de contaminación en estudio, las estimaciones de los costos que se incluyeron en la tabla 3.14 resultan inapropiadas y cada una puede estar equivocada hasta en 10% en cualquier dirección. También existe incertidumbre sobre los valores de los parámetros dados en la tabla 3.13, pero menos que sobre aquellos que se proporcionaron en la tabla 3.14. Por el contrario, los valores de la tabla 3.12 son estándares y, por tanto, son constantes preestablecidas.

Sin embargo, todavía existen diferentes puntos de vista acerca del nivel en que se deben establecer estos estándares sobre las reducciones requeridas de las tasas de emisión de los distintos contaminantes. En realidad, las cifras de la tabla 3.12 son valores preliminares tentativos que se acordaron antes de conocer los costos totales en los que se incurrirá para cumplirlos. Tanto las autoridades de la ciudad como los funcionarios de la compañía están de acuerdo en que la decisión final sobre los estándares debe basarse en un *equilibrio* entre *los* costos y los beneficios. Con esto en mente, las autoridades han concluido que cada incremento de 10% sobre los valores actuales de los estándares (los números de la tabla 3.12) tendrá un valor de $3.5 millones de dólares para la ciudad. Por ello, las autoridades acordaron reducir el pago de impuestos de la compañía en $3.5 millones de dólares por *cada* 10% de reducción de los estándares (hasta 50%) que ésta haya aceptado.

Por último, se analizaron los valores *relativos* de los estándares para los tres contaminantes. Como se indica en la tabla 3.12, por ahora la reducción requerida de las partículas es menos de la mitad de la que se fijó para los óxidos de azufre o los hidrocarburos. Algunas personas piensan que debe reducirse esta discrepancia. Otras sostienen que se justifica una diferencia aún mayor, pues los óxidos de sulfuro y los hidrocarburos causan un daño mucho mayor que las partículas. Se llegó a la conclusión de que este asunto se volverá a examinar después de obtener información sobre los trueques disponibles entre los estándares (aumento de uno y disminución de otro) sin que aumente el costo total.

***a*)** Utilice algún software de programación lineal para resolver el modelo que se formuló en la sección 3.4 para este problema. Además de la solución óptima, observe los resultados adicionales que proporcionó el análisis postóptimo (como el informe de sensibilidad al usar Excel), que constituyen la base de los siguientes pasos.

***b*)** Pase por alto las restricciones que no tienen incertidumbre respecto de sus parámetros (es decir, $x_j \leq 1$ para $j = 1, 2, \ldots, 6$), identifique los parámetros del modelo que deben clasificarse como *sensibles*. (*Sugerencia:* Vea la subsección "Análisis de sensibilidad" en la sección 4.7.) Si es posible, presente las recomendaciones pertinentes sobre qué parámetros deben tener una mejor estimación.

***c*)** Analice el efecto de la falta de precisión en la estimación de cada parámetro de costo que se presenta en la tabla 3.14. Si el valor verdadero es 10% *menor* que el valor estimado, ¿alterará este cambio la solución óptima? ¿Cambiaría ésta si el valor verdadero fuera 10% *mayor* que el valor estimado? Recomiende cuál debe ser el centro de atención para realizar una mejor estimación de los parámetros de costo.

***d*)** Considere el caso en que el modelo se convirtió a la forma de maximización antes de aplicar el método símplex. Utilice la tabla 6.14 para construir el problema dual correspondiente y use los resultados de aplicar el método símplex al problema primal para identificar una solución óptima para este problema dual. Si el primal se hubiera dejado en forma de minimización, ¿cómo se hubiera afectado a la forma del dual y al signo de las variables duales óptimas?

***e*)** Para cada contaminante, utilice los resultados del inciso *d*) para especificar la tasa de cambio del costo total de una solución óptima con cualquier cambio pequeño en la reducción requerida de la tasa anual de emisión del contaminante. Especifique también cuánto puede cambiar la reducción requerida (hacia arriba o hacia abajo) sin afectar la tasa de cambio del costo total.

***f*)** Por cada unidad de cambio en el estándar para las partículas que se presentó en la tabla 3.12, determine el cambio en la dirección opuesta para los óxidos de sulfuro que mantendría sin cambio el costo total de una solución óptima. Repita este procedimiento para los hidrocarburos. Después, hágalo en el caso de un cambio simultáneo e igual de los óxidos de sulfuro y de los hidrocarburos en dirección opuesta al de las partículas.

***g*)** Si θ denota el incremento porcentual de todos los estándares que se presentaron en la tabla 3.12, formule el problema de analizar el efecto de incrementos proporcionales simultáneos en estos estándares como un problema de programación lineal paramétrica. Después utilice los resultados del inciso *e*) para determinar la tasa a la que aumenta el costo total de una solución óptima con un cambio pequeño en el valor de θ a partir de cero.

***h*)** Utilice el método símplex para encontrar una solución óptima al problema de programación lineal paramétrica formulado en el inciso *g*) para cada $\theta = 10, 20, 30, 40, 50$. Considere el incentivo de impuestos ofrecido por las autoridades de la ciudad, y utilice

estos resultados para determinar el valor de θ (incluya la opción de $\theta = 0$) que debe seleccionar la compañía para minimizar el costo total tanto de la reducción de la contaminación como de los impuestos.

i) Para el valor θ seleccionado en el inciso *h*), repita los incisos *e*) y *f*) de manera que los tomadores de decisiones puedan decidir sobre los valores relativos de los estándares de los tres contaminantes.

RESUMEN DE LOS CASOS ADICIONALES EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

CASO 6.2 Administración de granjas

La familia Ploughman es dueña y opera una granja de 640 hectáreas que le ha pertenecido por varias generaciones. Ahora la familia necesita tomar una decisión acerca de la mezcla de ganado y cultivos para el próximo año. Si se supone que en el próximo año prevalecerán condiciones climáticas normales, se puede formular y resolver un modelo de programación lineal que sirva como guía al tomar esta decisión. Sin embargo, condiciones adversas del clima podrían dañar los cultivos y, en gran medida, reducir el valor resultante. Por tanto, se requiere de un análisis postóptimo que considere explorar el efecto de algunos escenarios posibles en el clima del próximo año y las implicaciones para la decisión de la familia.

CASO 6.3 Asignación de estudiantes a escuelas (revisado)

Ésta es una continuación del caso 4.3 en el que se involucra a la Oficina Escolar de Springfield, que debe asignar estudiantes provenientes de seis áreas residenciales en las tres escuelas de educación media que aún tienen plazas en la ciudad. Después de resolver un modelo del problema con algún paquete de software, el informe del análisis de sensibilidad de ese paquete necesita ser utilizado con dos propósitos. Uno es verificar el efecto de un aumento de ciertos costos de transportación por las obras en construcción de una de las áreas residenciales. El otro es explorar la posibilidad de agregar aulas portátiles para incrementar la capacidad de una o más de las escuelas por unos cuantos años.

CASO 6.4 Redacción de una síntesis ejecutiva

Después de establecer las metas sobre cuánto se deberían incrementar las ventas de tres productos como un resultado de una campaña de publicidad, la administración de la Profit & Gambit Co., ahora quiere explorar el trueque entre el costo de la publicidad y el incremento de las ventas. La primera tarea consiste en realizar el análisis de sensibilidad asociado. La tarea principal es elaborar una síntesis ejecutiva para la administración Profit & Gambit que presente los resultados en lenguaje administrativo.

CAPÍTULO 7

Otros algoritmos para programación lineal

La principal razón de que la programación lineal se utilice en forma tan amplia es la disponibilidad de un algoritmo excepcionalmente eficiente —el método símplex— que en forma rutinaria resuelve problemas grandes que con frecuencia surgen en la práctica. Sin embargo, el método símplex es sólo una parte del arsenal de algoritmos que se usan con regularidad en la aplicación de la programación lineal. Ahora se verán esos otros algoritmos.

Este capítulo comienza con tres algoritmos que son, en realidad, variantes del método símplex. En especial, las tres secciones siguientes presentan el *método símplex dual* (una modificación muy útil del análisis de sensibilidad), la *programación lineal paramétrica* (una extensión del análisis de sensibilidad sistemático) y la *técnica de ramificación y acotamiento* (una versión simplificada del método símplex para manejar variables que tienen cotas superiores).

En la sección 4.9 se introdujo un nuevo enfoque algorítmico para programación lineal, un tipo de algoritmo que se mueve por el interior de la región factible. En la sección 7.4 se describirá con detalle este *enfoque de punto interior*.

En un complemento de este capítulo, en el sitio en internet de este libro, también se introduce la *programación lineal por objetivos*, en la que, en lugar de tener *un solo objetivo* (maximizar o minimizar Z) como en programación lineal, el problema tiene *varios objetivos* que deben tratar de alcanzarse de manera simultánea. Ciertas técnicas de formulación permiten convertir un problema de programación lineal por objetivos en uno de programación lineal, de manera que se puedan usar los procedimientos de solución basados en el método símplex. En el complemento se describen estas técnicas y procedimientos.

7.1 MÉTODO SÍMPLEX DUAL

El *método símplex dual* se basa en la teoría de la dualidad que se presentó en la primera parte del capítulo 6. Para explicar la idea básica de este método, es útil usar parte de la terminología introducida en las tablas 6.10 y 6.11 de la sección 6.3 para describir un par de soluciones básicas complementarias de los problemas primal y dual. En particular, recuerde que se dice que ambas soluciones son *factibles primales* si la solución básica primal es factible, y se llaman *factibles duales* si la solución básica dual complementaria es factible para el problema dual. Recuerde también que —de acuerdo con el lado derecho de la tabla 6.11— cada solución básica complementaria es óptima para su problema sólo si es factible en el primal y en el dual.

Se puede pensar en el método símplex dual como un *reflejo* del método símplex. Éste trata directamente con soluciones básicas del problema primal que son *factibles primales* pero no factibles duales. Después se mueve hacia la solución óptima con el objetivo de lograr la factibilidad dual, la prueba de optimalidad del método símplex. Por el contrario, el método símplex dual maneja soluciones básicas del problema dual que son *factibles duales* pero no factibles primales. Después se mueve hacia la solución óptima con la meta de alcanzar la factibilidad en el primal.

Aún más, el método símplex dual trata un problema como si se aplicara el método símplex a su problema dual de manera simultánea. Si sus soluciones básicas *iniciales* se hacen *complementarias*, los dos métodos se mueven en una secuencia completa y obtienen soluciones básicas *complementarias* en cada iteración.

El método símplex dual es muy útil en algunas situaciones especiales. Lo normal es que sea más fácil encontrar una solución inicial básica que sea factible que una que sea factible en el dual. Sin embargo, en ocasiones es necesario introducir muchas variables *artificiales* para construir una solución inicial BF de manera artificial. En esos casos sería más sencillo comenzar con una solución básica factible dual y aplicar el método símplex dual. Incluso, puede ser que ésta requiera menos iteraciones cuando no se tiene que convertir en cero el valor de tantas variables artificiales.

Cuando se aborda un problema cuyas soluciones iniciales básicas —sin variables artificiales— no son factibles primales ni factibles duales, también es posible combinar las ideas de los métodos símplex y dual símplex en un algoritmo *primal-dual* que trate de alcanzar la factibilidad tanto primal como dual.

Como se mencionó varias veces en el capítulo 6 y en la sección 4.7, otra aplicación primordial del método símplex dual es su asociación con el análisis de sensibilidad. Suponga que se tiene una solución óptima por el método símplex, pero que es necesario —o de interés para el análisis de sensibilidad— introducir cambios menores en el modelo. Si la solución básica óptima que se tenía *ya no es factible primal* —pero todavía satisface la prueba de optimalidad—, se puede aplicar de inmediato el método símplex dual a partir de esta solución básica *factible dual* (al final de esta sección se mostrará este aspecto). Por lo general, mediante la aplicación del método símplex dual se llega a la nueva solución óptima mucho más rápido que si se resuelve el nuevo problema desde el principio con el método símplex.

El método símplex dual también puede ser útil para resolver problemas grandes de programación lineal desde el principio debido a la eficiencia del algoritmo. La experiencia computacional reciente con las últimas versiones de CPLEX indica que el método símplex dual suele ser más eficiente que el método símplex para resolver problemas particularmente grandes que se encuentran en la práctica.

Las reglas del método símplex dual son muy parecidas a las del método símplex. En realidad, una vez que los métodos se inician, la única diferencia entre ellos es el criterio para elegir las variables básicas que entran y salen y la regla para detener el algoritmo.

Para poner en marcha el método símplex dual —si se trata de un problema de maximización—, todos los coeficientes de la ecuación (0) deben ser *no negativos* —de manera que la solución básica es factible dual—. Las soluciones básicas serán no factibles —excepto la última— sólo porque algunas variables son negativas. El método hace que el valor de la función objetivo disminuya de manera continua, y conserve siempre *coeficientes no negativos* en la ecuación (0), hasta que todas las *variables* sean no negativas. En este momento la solución básica es factible —satisface todas las ecuaciones— y, por tanto, es óptima según el criterio del método símplex de coeficientes no negativos de la ecuación (0).

A continuación se resumen los detalles del método símplex dual.

Resumen del método símplex dual

1. *Paso inicial:* Después de convertir cualquier restricción funcional de la forma $\geq$ a la forma $\leq$ —mediante la multiplicación de ambos lados por -1—, se introducen las variables de holgura necesarias para construir un conjunto de ecuaciones que describan el problema. Encuentre una solución básica tal que los coeficientes de la ecuación (0) sean ceros para el caso de las variables básicas y no negativos para las variables no básicas (de manera que la solución es óptima si es factible). El método continúa en la prueba de factibilidad.

2. *Prueba de factibilidad:* Se verifica si todas las variables básicas son *no negativas*. Si es así, esta solución es factible y, por tanto, óptima, y el algoritmo se detiene. De otra manera, es necesario realizar una iteración.

3. *Iteración:*

Paso 1 Se determina la *variable básica que sale* mediante la selección de la variable básica negativa que tenga el mayor valor absoluto.

Paso 2 Se determina la *variable básica entrante* a través de la selección de aquella cuyo coeficiente en la ecuación (0) llegue primero a cero al agregar a la ecuación (0) un múltiplo

creciente de la ecuación que contiene a la variable básica que sale. Para llevar a cabo esta elección se examinan las variables no básicas con *coeficientes negativos* de esa ecuación —la que contiene la variable básica que sale— y se elige la que tiene el menor valor absoluto del cociente dado por el coeficiente de la ecuación (0) entre el coeficiente de esa ecuación.

Paso 3 Se determina la *nueva solución básica* a partir del conjunto actual de ecuaciones y se despejan las variables básicas en términos de las no básicas, mediante el método de eliminación gaussiana. Cuando las variables no básicas se hacen cero, cada variable básica (y Z) es igual al nuevo valor del lado derecho de la ecuación en la que aparece (con coeficiente $+1$). Se repite la prueba de factibilidad.

Para comprender bien el método símplex dual se debe entender que hace lo mismo que haría el *método símplex* si se aplicara a las soluciones básicas complementarias del *problema dual* (en realidad, éste fue el motivo para construirlo así). El paso 1 de una iteración, que determina la variable básica que sale, es equivalente a determinar la variable básica entrante en el problema dual. La variable negativa con el valor absoluto más grande corresponde al coeficiente negativo con el mayor valor absoluto de la ecuación (0) del problema dual (vea la tabla 6.3). El paso 2, que determina la variable básica entrante, es equivalente a determinar la variable básica que sale en el problema dual. El coeficiente de la ecuación (0) que llega primero a cero corresponde a la primera variable que se hace cero en el problema dual. También los dos criterios para detener el algoritmo son complementarios.

Un ejemplo

Ahora se ilustrará el método símplex dual al aplicarlo al *problema dual* de la Wyndor Glass Co. (vea la tabla 6.1). Por lo general, este método se aplica de manera directa al problema en cuestión (un problema primal). Sin embargo, se escogió este problema porque ya se conoce la aplicación del método símplex a su problema dual —es decir, al problema primal[1]—, la cual se muestra en la tabla 4.8, de tal forma que se pueden comparar. Para facilitar la comparación, las variables de decisión del problema que se va a resolver se denotarán como y_i en lugar de x_j.

En forma de *maximización* el problema que se pretende resolver es

$$\text{Maximizar} \quad Z = -4y_1 - 12y_2 - 18y_3,$$

sujeta a

$$\begin{aligned} y_1 \qquad + 3y_3 &\geq 3 \\ 2y_2 + 2y_3 &\geq 5 \end{aligned}$$

y

$$y_1 \geq 0, \qquad y_2 \geq 0, \qquad y_3 \geq 0.$$

Como ahora se permiten valores negativos en el lado derecho, no es necesario introducir variables artificiales como variables básicas iniciales. En su lugar, sólo se convierten las restricciones funcionales a la forma y se introducen las variables de holgura para que desempeñen este papel. El conjunto inicial de ecuaciones se muestra en la iteración 0 de la tabla 7.1. Observe que todos los coeficientes de la ecuación (0) son no negativos, con lo que la solución es óptima si es que resulta factible.

La solución básica inicial es $y_1 = 0$, $y_2 = 0$, $y_3 = 0$, $y_4 = -3$, $y_5 = -5$, con $Z = 0$, que no es factible porque tiene valores negativos. La variable básica que sale es y_5 ($5 > 3$), y la variable básica entrante es y_2 ($12/2 < 18/2$), lo que conduce al segundo conjunto de ecuaciones que se muestra como la iteración 1 de la tabla 7.1. La solución básica correspondiente es $y_1 = 0$, $y_2 = \frac{5}{2}$, $y_3 = 0$, $y_4 = -3$, $y_5 = 0$, con $Z = -30$, que no es factible.

La siguiente variable básica que sale es y_4, y la que entra es y_3 ($6/3 < 4/1$), lo que produce el conjunto final de ecuaciones que se presenta en la tabla 7.1. La solución básica correspondiente es $y_1 = 0$, $y_2 = \frac{3}{2}$, $y_3 = 1$, $y_4 = 0$, $y_5 = 0$, con $Z = -36$, que es factible y por tanto óptima.

[1] Recuerde que la propiedad de simetría que se explicó en la sección 6.1 indica que el dual de un problema dual es el problema primal original.

TABLA 7.1 Método símplex dual aplicado al problema de la Wyndor Glass Co.

Iteración	Variable básica	Ec.	Coeficiente de: Z	y_1	y_2	y_3	y_4	y_5	Lado derecho
0	Z	(0)	1	4	12	18	0	0	0
	y_4	(1)	0	−1	0	−3	1	0	−3
	y_5	(2)	0	0	−2	−2	0	1	−5
1	Z	(0)	1	4	0	6	0	6	−30
	y_4	(1)	0	−1	0	−3	1	0	−3
	y_2	(2)	0	0	1	1	0	$-\frac{1}{2}$	$\frac{5}{2}$
2	Z	(0)	1	2	0	0	2	6	−36
	y_3	(1)	0	$\frac{1}{3}$	0	1	$-\frac{1}{3}$	0	1
	y_2	(2)	0	$-\frac{1}{3}$	1	0	$\frac{1}{3}$	$-\frac{1}{2}$	$\frac{3}{2}$

Observe que la solución óptima para el dual de este problema[2] es $x_1^* = 2, x_2^* = 6, x_3^* = 2, x_4^* = 0$, $x_5^* = 0$, que es la misma que se obtuvo en la tabla 4.8 mediante el método símplex. Se sugiere al lector que siga con detalle al mismo tiempo lo que se obtuvo en las tablas 7.1 y 4.8 y que compare los pasos complementarios de los dos métodos con imágenes de espejo o simétricas.

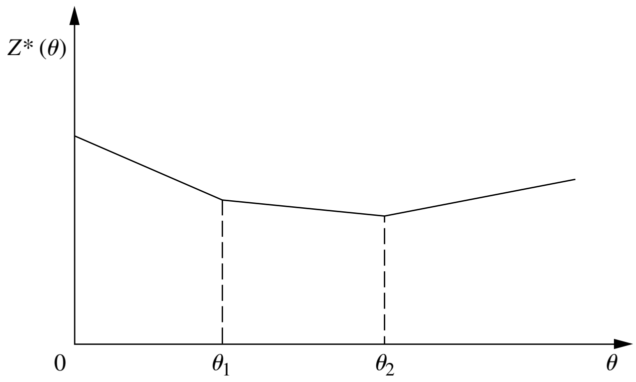
Como se mencionó anteriormente, una aplicación de primordial importancia del método símplex dual consiste en que con frecuencia éste puede utilizarse para resolver un problema de manera rápida cuando el análisis de sensibilidad consista en hacer pequeños cambios al modelo original. En particular, si la solución básica óptima inicial no es factible primal (uno o más coeficientes del lado derecho son negativos) pero aún satisfacen la prueba de optimalidad (el renglón 0 no tiene coeficientes negativos), usted puede aplicar de inmediato el método símplex dual comenzando con esta solución dual básica factible. Por ejemplo, esta situación surge cuando una nueva restricción que viola la solución óptima inicial se agrega al modelo original. Para ilustrar lo anterior, suponga que el problema que se resolvió en la tabla 7.1 no incluía su primera restricción funcional ($y_1 + 3y_3 \geq 3$). Después de eliminar el renglón 1, la iteración de la tabla 1 de la tabla 7.1 muestra que la solución óptima resultante es $y_1 = 0$, $y_2 = \frac{5}{2}$, $y_3 = 0$, $y_5 = 0$, con $Z = -30$. Ahora suponga que el análisis de sensibilidad lo conduce a sumar la restricción que originalmente se omitió, $y_1 + 3y_3 \geq 3$, la cual es violada por la solución óptima original puesto que $y_1 = 0$ y $y_3 = 0$. Para encontrar la nueva solución óptima, esta restricción (incluyendo la variable de huelgo y_4) podría sumarse como renglón 1 de la tabla de en medio en la tabla 7.1. Sin tomar en cuenta si esta tabla se obtuvo mediante la aplicación del método símplex o el símplex dual para obtener la solución óptima original (quizá después de muchas iteraciones), la aplicación del método símplex dual a esta tabla nos lleva a la nueva solución óptima en una sola iteración.

Si desea ver **otro ejemplo** de aplicación del método símplex dual, ingrese a la sección de Worked Examples en la página en internet de este libro.

7.2 PROGRAMACIÓN LINEAL PARAMÉTRICA

Al final de la sección 6.7 se describió la *programación lineal paramétrica* y su utilización para realizar un análisis de sensibilidad sistemático que cambia en forma gradual y simultánea varios parámetros del modelo. A continuación se presentará el procedimiento algorítmico, primero para el caso en el que se cambian los parámetros c_j y después cuando se modifican los parámetros b_i.

[2] La *propiedad de soluciones básicas óptimas complementarias* que se presentó en la sección 6.3 indica cómo leer la solución óptima del problema dual del renglón 0 de la tabla símplex final del problema primal. Esta misma conclusión es válida sin que importe si se aplicó el método símplex, o el símplex dual, para obtener la tabla final.



■ **FIGURA 7.1**
Valor de la función objetivo de una solución óptima como una función de θ para programación lineal paramétrica con cambios sistemáticos en los parámetros c_j.

Cambios sistemáticos en los parámetros c_j

En el caso en el que se cambian los parámetros c_j la *función objetivo* del modelo de programación lineal normal

$$Z = \sum_{j=1}^{n} c_j x_j$$

se sustituye por

$$Z(\theta) = \sum_{j=1}^{n} (c_{j+}\alpha_j\theta)x_j,$$

donde las α_j son constantes de entrada dadas que representan las tasas *relativas* a las que cambian los coeficientes. Al incrementar θ en forma gradual desde cero, los coeficientes cambian a estas tasas relativas.

Los valores asignados a las α_j pueden representar cambios simultáneos interesantes de las c_j para realizar un análisis de sensibilidad sistemático del efecto que tiene un aumento de la magnitud de estos cambios. También pueden estar basados en la forma como cambiarían los coeficientes (por ejemplo, ganancias unitarias) respecto de algún factor medido por θ. Este factor puede ser incontrolable, como el estado de la economía. No obstante, también puede encontrarse bajo el control del tomador de decisiones, por ejemplo, la cantidad de personal y equipo que debe cambiarse de una actividad a otra.

Para cualquier valor dado de θ, se puede obtener la solución óptima del problema de programación lineal correspondiente, a través del método símplex. Tal vez esta solución ya se obtuvo para el problema original, en donde $\theta = 0$. Sin embargo, el objetivo es *encontrar la solución óptima* del problema modificado de programación lineal [maximizar $Z(\theta)$ sujeta a las restricciones originales] como *una función* de θ. Así, es necesario que el procedimiento de solución sea capaz de determinar cómo y cuándo cambia la solución óptima (si lo hace) cuando el valor de θ aumenta de cero a cualquier número positivo específico.

En la figura 7.1 se representa la forma en que el valor de la función objetivo $Z^*(\theta)$, de la solución óptima, cambia al incrementarse el valor de θ. En realidad, $Z^*(\theta)$ siempre tiene esta forma lineal *por partes* y *convexa*[3] (vea el problema 7.2-7). La solución óptima correspondiente cambia (al aumentar θ) *exactamente* en los valores de θ donde cambia la pendiente de la función $Z^*(\theta)$. Entonces, la figura 7.1 ilustra un problema en donde tres soluciones diferentes son óptimas para tres valores diferentes de θ, la primera para $0 \leq \theta \leq \theta_1$, la segunda para $\theta_1 \leq \theta \leq \theta_2$ y la tercera para $\theta \geq \theta_2$. Como el valor de cada x_j permanece igual dentro de cada uno de estos intervalos de θ, el valor de $Z^*(\theta)$ varía sólo con θ porque los *coeficientes* de x_j cambian como una función lineal de θ. Este procedimiento de solución se basa de manera directa en el procedimiento de análisis de sensibilidad para investigar los cambios en los parámetros c_j (casos 2*a* y 3, sección 6.7). Como ya

[3] En el apéndice 2 se da una definición y se examinan las funciones convexas.

se describió en la última parte de la sección 6.7, la única diferencia básica es que ahora los cambios se expresan en términos de θ y no de números específicos.

Ejemplo. Para ilustrar el procedimiento de solución, suponga que $\alpha_1 = 2$ y $\alpha_2 = -1$ para el problema original de la Wyndor Glass Co. (véase la sección 3.1), de manera que

$$Z(\theta) = (3 + 2\theta)x_1 + (5 - \theta)x_2.$$

Si se inicia con la tabla símplex final para $\theta = 0$ (tabla 4.8), se ve que su ecuación (0)

$$(0) \qquad Z + \frac{3}{2}x_4 + x_5 = 36$$

tendría primero los siguientes cambios cuando se agreguen los coeficientes originales ($\theta = 0$) en el lado izquierdo:

$$(0) \qquad Z - 2\theta x_1 + \theta x_2 + \frac{3}{2}x_4 + x_5 = 36.$$

Dado que x_1 y x_2 son varibles básicas [que aparecen en las ecuaciones (3) y (2), respectivamente], es necesario eliminarlas algebraicamente de la ecuación (0):

$$\begin{array}{rl} & Z - 2\theta x_1 + \theta x_2 + \frac{3}{2}x_4 + x_5 = 36 \\ & \quad + \ 2\theta \text{ multiplica la ec. (3)} \\ & \quad - \ \ \theta \text{ multiplica la ec. (2)} \\ \hline (0) & Z + \left(\frac{3}{2} - \frac{7}{6}\theta\right)x_4 + \left(1 + \frac{2}{3}\theta\right)x_5 = 36 - 2\theta. \end{array}$$

La prueba de optimalidad afirma que la solución básica factible actual permanecerá óptima mientras estos coeficientes de las variables no básicas sean no negativos:

TABLA 7.2 Procedimiento de programación lineal del parámetro c_j aplicado al ejemplo de la Wyndor Glass Co.

Intervalo de θ	Variable básica	Ec.	Coeficiente de: Z	x_1	x_2	x_3	x_4	x_5	Lado derecho	Solución óptima
$0 \le \theta \le \frac{9}{7}$	$Z(\theta)$	(0)	1	0	0	0	$\frac{9 - 7\theta}{6}$	$\frac{3 + 2\theta}{3}$	$36 - 2\theta$	$x_4 = 0$ $x_5 = 0$
	x_3	(1)	0	0	0	1	$\frac{1}{3}$	$-\frac{1}{3}$	2	$x_3 = 2$
	x_2	(2)	0	0	1	0	$\frac{1}{2}$	0	6	$x_2 = 6$
	x_1	(3)	0	1	0	0	$-\frac{1}{3}$	$\frac{1}{3}$	2	$x_1 = 2$
$\frac{9}{7} \le \theta \le 5$	$Z(\theta)$	(0)	1	0	0	$\frac{-9 + 7\theta}{2}$	0	$\frac{5 - \theta}{2}$	$27 + 5\theta$	$x_3 = 0$ $x_5 = 0$
	x_4	(1)	0	0	0	3	1	-1	6	$x_4 = 6$
	x_2	(2)	0	0	1	$-\frac{3}{2}$	0	$\frac{1}{2}$	3	$x_2 = 3$
	x_1	(3)	0	1	0	1	0	0	4	$x_1 = 4$
$\theta \ge 5$	$Z(\theta)$	(0)	1	0	$-5 + \theta$	$3 + 2\theta$	0	0	$12 + 8\theta$	$x_2 = 0$ $x_3 = 0$
	x_4	(1)	0	0	2	0	1	0	12	$x_4 = 12$
	x_5	(2)	0	0	2	-3	0	1	6	$x_5 = 6$
	x_1	(3)	0	1	0	1	0	0	4	$x_1 = 4$

$$\frac{3}{2} - \frac{7}{6}\theta \geq 0, \qquad \text{para } 0 \leq \theta \leq \frac{9}{7},$$

$$1 + \frac{2}{3}\theta \geq 0, \qquad \text{para toda } \theta \geq 0.$$

Entonces, si se aumenta θ a un valor más alto que $\theta = \frac{9}{7}$, x_4 se convierte en la variable básica entrante en la siguiente iteración del método símplex, para encontrar la nueva solución óptima. Después θ aumenta más hasta encontrar otro coeficiente que se hace negativo, y así sucesivamente, hasta que θ logra el incremento deseado.

A continuación se resumirá el procedimiento completo y se completará el ejemplo en la tabla 7.2.

Resumen del procedimiento de programación lineal paramétrica para realizar cambios sistemáticos en los parámetros c_j

1. Resuelva el problema con $\theta = 0$ por medio del método símplex.
2. Utilice el procedimiento de análisis de sensibilidad (casos 2*a* y 3, sección 6.7) para introducir los cambios $\Delta c_j = \alpha_j \theta$ en la ecuación (0).
3. Incremente θ hasta que el coeficiente de una de las variables no básicas de la ecuación (0) se vuelva negativo (o hasta incrementar todo lo que se desea).
4. Use esta variable como variable básica entrante para llevar a cabo una nueva iteración del método símplex, para encontrar una nueva solución óptima. Regrese al paso 3.

Cambios sistemáticos en los parámetros b_i

En el caso en que se hacen cambios sistemáticos en los parámetros b_i, la única modificación que se introduce al modelo de programación lineal es sustituir b_i por $b_i + \alpha_i\theta$, para $i = 1, 2, \ldots, m$, donde las α_i son constantes de entrada dadas. Así, el problema se convierte en

$$\text{Maximizar} \qquad Z(\theta) = \sum_{j=1}^{n} c_j x_j,$$

sujeta a

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i + \alpha_i \theta \qquad \text{para } i = 1, 2, \ldots, m$$

y

$$x_j \geq 0 \qquad \text{para } j = 1, 2, \ldots, n.$$

La meta es identificar la solución óptima como una función de θ.

De acuerdo con esta formulación, el valor correspondiente de la función objetivo $Z^*(\theta)$ siempre tiene la *forma lineal por partes* y *cóncava*[4] que se muestra en la figura 7.2 (vea el problema 7.2-8). El conjunto de variables básicas de la solución óptima cambia (al aumentar θ) *sólo* en el punto en que cambia la pendiente de $Z^*(\theta)$, pero ahora, al contrario del caso anterior, los valores de estas variables cambian como una función (lineal) de θ entre los cambios de pendiente. La razón es que al incrementar θ cambian los valores de los coeficientes del lado derecho del conjunto inicial de ecuaciones, lo que a su vez ocasiona cambios en los valores del lado derecho del conjunto final de ecuaciones, es decir, en los valores del conjunto final de variables básicas. En la figura 7.2 se hace un bosquejo de un problema con tres conjuntos de variables básicas que son óptimas para valores diferentes de θ, uno para $0 \leq \theta \leq \theta_1$, el segundo para $\theta_1 \leq \theta \leq \theta_2$ y el tercero para $\theta \geq \theta_2$. Dentro de cada uno de estos intervalos de θ, el valor de $Z^*(\theta)$ varía con θ a pesar de los coeficientes fijos c_j porque los valores de las x_j cambian.

[4] En el apéndice 2 se da una definición y se examinan las funciones cóncavas.

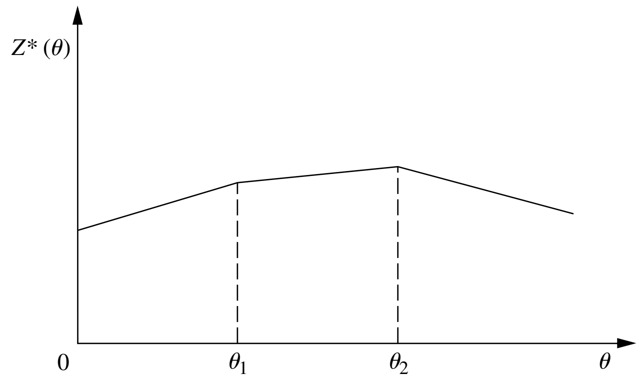


FIGURA 7.2
Valor de la función objetivo de una solución óptima como una función de θ para programación lineal paramétrica con cambios sistemáticos en los parámetros b_i.

El siguiente resumen del procedimiento de solución es muy parecido al que se acaba de presentar respecto a los cambios sistemáticos en los parámetros c_j. Esto se debe a que cambiar las b_i es equivalente a cambiar los coeficientes de la función objetivo del modelo *dual*. Entonces, el procedimiento aplicable al problema primal es exactamente *complementario* a la aplicación simultánea del procedimiento para cambios sistemáticos en los parámetros c_j del problema *dual*. En consecuencia, ahora se usa el *método símplex dual* (vea la sección 7.1) para obtener cada nueva solución óptima, y el caso de la aplicación del análisis de sensibilidad (vea la sección 6.7) ahora es el caso 1, pero estas diferencias son las únicas importantes.

Resumen del procedimiento de programación lineal paramétrica para cambios sistemáticos en los parámetros b_i

1. Resuelva el problema con $\theta = 0$ por medio del método símplex.
2. Utilice el procedimiento de análisis de sensibilidad (caso 1, sección 6.7) para introducir los cambios $\Delta b_i = \alpha_i \theta$ en la columna del *lado derecho*.
3. Incremente θ hasta que el valor de una de las variables básicas de la columna del *lado derecho* se vuelva negativo (o hasta incrementar θ todo lo que se desea).
4. Use esta variable como variable básica que sale para llevar a cabo una nueva iteración del método dual símplex, para encontrar una nueva solución óptima. Regrese al paso 3.

Ejemplo. Este procedimiento se aplicará al problema dual de la Wyndor Glass Co. (vea la tabla 6.1) con el fin de resaltar su relación de *dualidad* con el procedimiento para manejar los cambios sistemáticos en los parámetros c_j. En particular, suponga que $\alpha_1 = 2$ y $\alpha_2 = -1$ con lo que las restricciones funcionales se convierten en

$$\begin{array}{llll} y_1 \quad + 3y_3 \geq 3 + 2\theta & \text{o} & -y_1 \quad - 3y_3 \leq -3 - 2\theta \\ \quad 2y_2 + 2y_3 \geq 5 - \theta & \text{o} & \quad -2y_2 - 2y_3 \leq -5 + \theta. \end{array}$$

Así, el dual de *este* problema es el ejemplo que se consideró en la tabla 7.2.

Este problema se resolvió en la tabla 7.1 con $\theta = 0$ por lo que se partirá de la tabla símplex final que se obtuvo. Si se emplea el procedimiento de análisis de sensibilidad para el caso 1, de la sección 6.7, se puede observar que los elementos de la columna del *lado derecho* de esta tabla cambian a los valores siguientes:

$$Z^* = \mathbf{y}^*\bar{\mathbf{b}} = [2, 6]\begin{bmatrix} -3 - 2\theta \\ -5 + \theta \end{bmatrix} = -36 + 2\theta,$$

$$\mathbf{b}^* = \mathbf{S}^*\bar{\mathbf{b}} = \begin{bmatrix} -\frac{1}{3} & 0 \\ \frac{1}{3} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} -3 - 2\theta \\ -5 + \theta \end{bmatrix} = \begin{bmatrix} 1 + \frac{2\theta}{3} \\ \frac{3}{2} - \frac{7\theta}{6} \end{bmatrix}.$$

TABLA 7.3 Procedimiento de programación paramétrica de b_i aplicado al dual del ejemplo de la Wyndor Glass Co.

Intervalo de θ	Variable básica	Ec.	Coeficiente de: Z	y_1	y_2	y_3	y_4	y_5	Lado derecho	Solución óptima
$0 \le \theta \le \frac{9}{7}$	$Z(\theta)$	(0)	1	2	0	0	2	6	$-36 + 2\theta$	$y_1 = y_4 = y_5 = 0$
	y_3	(1)	0	$\frac{1}{3}$	0	1	$-\frac{1}{3}$	0	$\frac{3 + 2\theta}{3}$	$y_3 = \frac{3 + 2\theta}{3}$
	y_2	(2)	0	$-\frac{1}{3}$	1	0	$\frac{1}{3}$	$-\frac{1}{2}$	$\frac{9 - 7\theta}{6}$	$y_2 = \frac{9 - 7\theta}{6}$
$\frac{9}{7} \le \theta \le 5$	$Z(\theta)$	(0)	1	0	6	0	4	3	$-27 - 5\theta$	$y_2 = y_4 = y_5 = 0$
	y_3	(1)	0	0	1	1	0	$-\frac{1}{2}$	$\frac{5 - \theta}{2}$	$y_3 = \frac{5 - \theta}{2}$
	y_1	(2)	0	1	-3	0	-1	$\frac{3}{2}$	$\frac{-9 + 7\theta}{2}$	$y_1 = \frac{-9 + 7\theta}{2}$
$\theta \ge 5$	$Z(\theta)$	(0)	1	0	12	6	4	0	$-12 - 8\theta$	$y_2 = y_3 = y_4 = 0$
	y_5	(1)	0	0	-2	-2	0	1	$-5 + \theta$	$y_5 = -5 + \theta$
	y_1	(2)	0	1	0	3	-1	0	$3 + 2\theta$	$y_1 = 3 + 2\theta$

Por tanto, las dos variables básicas de esta tabla símplex,

$$y_3 = \frac{3 + 2\theta}{3} \qquad \text{y} \qquad y_2 = \frac{9 - 7\theta}{6}$$

permanecen no negativas para $0 \le \theta \le \frac{9}{7}$. Si adquiere un valor θ mayor que $\theta = \frac{9}{7}$ es necesario que y_2 se convierta en una variable básica que sale para realizar otra iteración del método símplex, y así sucesivamente, tal como se resume en la tabla 7.3.

Se sugiere al lector que siga los cálculos de las tablas 7.2 y 7.3 al mismo tiempo que observa la relación de dualidad entre los dos procedimientos.

En la sección Worked Examples del sitio en internet de este libro se incluye **otro ejemplo** del procedimiento para cambios sistemáticos en los parámetros b_i.

7.3 TÉCNICA DE LA COTA SUPERIOR

En problemas de programación lineal es común que algunas o todas las variables x_j *individuales* tengan *restricciones de cota superior*

$$x_j \le u_j,$$

donde u_j es una constante positiva que representa el máximo valor *factible* de x_j. En la sección 4.8 se puntualizó que el factor determinante en cuanto al tiempo de cálculo cuando se aplica el método símplex es el *número de restricciones funcionales*, mientras que el número de restricciones de *no negatividad* casi carece de importancia. En consecuencia, un número grande de restricciones de cota superior incluidas en las restricciones funcionales incrementa en gran medida el esfuerzo computacional requerido.

La *técnica de la cota superior* evita este mayor esfuerzo pues elimina las restricciones de cota superior del conjunto de restricciones funcionales y las trata por separado, en esencia, como restricciones de no negatividad.[5] Utilizar este procedimiento no causa problemas siempre y cuando

[5] La técnica de cota superior supone que las variables tienen las restricciones usuales de no negatividad además de las restricciones de cota superior. Si una variable tiene una cota inferior distinta a 0, por ejemplo, $x_j \ge L_j$, entonces esta restricción se puede convertir en una de no negatividad al hacer el cambio de variable, $x_j' = x_j - L_j$, de manera que $x_j' \ge 0$.

ninguna de las variables adquiera un valor mayor que su cota superior. La única vez que el método símplex incrementa alguna de las variables es cuando la variable básica entrante aumenta su valor para obtener la nueva solución BF. La técnica de la cota superior sólo aplica el método símplex al *resto* del problema —es decir, sin las restricciones de cota superior—, pero con la restricción adicional de que cada nueva solución básica factible debe satisfacer las restricciones de cota superior además de las normales de cota inferior (no negatividad).

Para aplicar esta idea, observe que una variable de decisión x_j con una restricción de cota superior $x_j \leq u_j$ siempre se puede sustituir por

$$x_j = u_j - y_j,$$

donde y_j será entonces la variable de decisión. En otras palabras, se puede elegir entre dos tipos de variables de decisión, la cantidad *mayor que cero* (x_j) o la *cantidad menor que* u_j ($y_j = u_j - x_j$). (Se hará referencia a x_j y y_j como variables de decisión *complementarias*.) Debido a que

$$0 \leq x_j \leq u_j$$

también se cumple que

$$0 \leq y_j \leq u_j.$$

Entonces, en cualquier momento, cuando se trabaja con el método símplex se puede

1. Usar x_j, donde $0 \leq x_j \leq u_j$, o
2. Sustituir x_j por $u_j - y_j$, donde $0 \leq y_j \leq u_j$.

La técnica de la cota superior emplea la siguiente regla para hacer esta elección:

Regla: Comience con la opción 1.
Si $x_j = 0$, use la opción 1, de manera que x_j es *no básica*.
Si $x_j = u_j$, use la opción 2, de manera que $y_j = 0$ es *no básica*.
Se cambia de opción sólo cuando x_j alcance el otro valor extremo.

Por tanto, siempre que una variable básica llega a su cota superior debe cambiar de opción y usar su variable de decisión complementaria como la nueva variable no básica (la variable que sale) para identificar la nueva solución BF. En consecuencia, la única modificación sustantiva que se hizo al método símplex radica en la regla para elegir la variable básica que sale.

Recuerde que el método símplex elige como variable básica que sale a aquella que sería la primera en convertirse en no factible al tomar valores negativos cuando se incrementa el valor de la variable básica entrante. En cambio, con la modificación que se acaba de hacer, se selecciona la variable que sería la primera en volverse no factible en *cualquier dirección*, ya sea por volverse negativa o por sobrepasar la cota superior cuando se incrementa la variable básica entrante. (Observe que una posibilidad es que la variable básica entrante se convierta en no factible si adquiere un valor mayor que su cota superior, caso en el cual su variable complementaria se convierte en la variable básica que sale.) Si la variable básica que sale adquiere el valor cero, se procede con el método símplex en forma normal, pero si por el contrario alcanza su cota superior, entonces se cambia de opción y su variable de decisión complementaria será la variable básica que sale.

Un ejemplo

Para ilustrar este procedimiento, considere el problema:

$$\text{Maximizar} \quad Z = 2x_1 + x_2 + 2x_3,$$

sujeta a

$$\begin{aligned} 4x_1 + x_2 &= 12 \\ -2x_1 + x_3 &= 4 \end{aligned}$$

y

$$0 \leq x_1 \leq 4, \qquad 0 \leq x_2 \leq 15, \qquad 0 \leq x_3 \leq 6.$$

TABLA 7.4 Ecuaciones y cálculos de la variable básica inicial que sale en el ejemplo de la técnica de cota superior

Grupo inicial de ecuaciones	Valor factible máximo de x_1
(0) $Z - 2x_1 \qquad\qquad = 20$	$x_1 \leq 4$ (desde $u_1 = 4$)
(1) $\quad 4x_1 + x_2 \qquad = 12$	$x_1 \leq \frac{12}{4} = 3$
(2) $\quad -2x_1 \qquad + x_3 = 4$	$x_1 \leq \frac{6-4}{2} = 1 \leftarrow$ mínimo (puesto que $u_3 = 6$)

Así, las tres variables tienen restricciones de cota superior ($u_1 = 4$, $u_2 = 15$, $u_3 = 6$).

Las dos restricciones de igualdad se encuentran ya en la forma apropiada de eliminación de Gauss para identificar la solución BF inicial ($x_1 = 0$, $x_2 = 12$, $x_3 = 4$), y ninguna de las variables de esta solución excede su cota superior; entonces, x_2 y x_3 se pueden usar como variables básicas iniciales sin introducir variables artificiales. De cualquier manera, es necesario eliminar algebraicamente estas variables de la función objetivo para obtener la ecuación (0) inicial, como se muestra a continuación:

$$\begin{array}{lrl} & Z - 2x_1 - x_2 - 2x_3 &= 0 \\ & +(4x_1 + x_2 \qquad\quad &= 12) \\ & +2(-2x_1 \qquad + x_3 &= 4) \\ \hline (0) & Z - 2x_1 \qquad\qquad\quad &= 20. \end{array}$$

Para comenzar con la primera iteración, esta ecuación (0) inicial indica que la variable básica *entrante* inicial es x_1. Como las restricciones de cota superior no deben incluirse, en la tabla 7.4 se muestra el conjunto inicial completo de ecuaciones y los cálculos correspondientes para seleccionar la variable básica que sale. La segunda columna indica cuánto puede *aumentar* la variable básica entrante x_1 desde cero y antes de que alguna variable básica (incluso x_1) se vuelva no factible. Ahora, el valor máximo que se da a la ecuación (0) es sólo la cota superior de x_1. En la ecuación (1), como el coeficiente de x_1 es *positivo*, al *aumentar* x_1 a 3 su valor, la variable básica (x_2) disminuye de 12 a su cota inferior de *cero*. En la ecuación (2), el coeficiente de x_1 es *negativo*, por lo que si se *aumenta* su valor a 1 la variable básica x_1 *aumenta* de 4 a su *cota superior* a 6.

Como la ecuación (2) contiene el valor factible *máximo* más *pequeño* de x_1 de la tabla 7.4, la variable básica de esta ecuación (x_3) es la variable básica que *sale*. Sin embargo, como x_3 alcanzó su cota *superior*, se sustituye por $6 - y_3$, y así $y_3 = 0$ se convierte en la nueva variable no básica de la siguiente solución BF y x_1 en la nueva variable básica de la ecuación (2). Este reemplazo conduce a los siguientes cambios en dicha ecuación:

$$\begin{array}{lrl} (2) & -2x_1 + x_3 &= 4 \\ & \rightarrow -2x_1 + 6 - y_3 &= 4 \\ & \rightarrow -2x_1 - y_3 &= -2 \\ & \rightarrow x_1 + \frac{1}{2}y_3 &= 1 \end{array}$$

Entonces, después de eliminar algebraicamente x_1 de todas las demás ecuaciones, el segundo conjunto completo de ecuaciones es

$$\begin{array}{lrl} (0) & Z \qquad + y_3 &= 22 \\ (1) & x_2 - 2y_3 &= 8 \\ (2) & x_1 \quad + \frac{1}{2}y_3 &= 1. \end{array}$$

La solución BF que se obtiene es $x_1 = 1$, $x_2 = 8$, $y_3 = 0$. De acuerdo con la prueba de optimalidad, se trata también de una solución óptima, por lo que $x_1 = 1$, $x_2 = 8$, $x_3 = 6 - y_3 = 6$ es la solución que se busca para el problema original.

Si se desea ver **otro ejemplo** de la técnica de la cota superior, en la sección Worked Examples en el sitio en internet de este libro se incluye uno.

7.4 ALGORITMO DE PUNTO INTERIOR

En la sección 4.9 se presentó un desarrollo muy importante en la programación lineal que tuvo lugar en 1984, y se debe a Narendra Karmarkar, de AT&T Bell Laboratories. Se trata de un poderoso algoritmo para resolver problemas muy grandes de programación lineal con una idea muy diferente a la que subyace al método símplex. Se introducirá aquí la naturaleza del enfoque de Karmarkar con la descripción de una variante relativamente sencilla —la variante "afín" o "escala afín"— de este algoritmo.[6] (En el IOR Tutorial también se incluye esta variante bajo el título *Solve Automatically by the Interior-Point Algorithm*.)

Para evitar los detalles matemáticos, esta sección se enfocará en las principales ideas de Karmarkar a un nivel intuitivo. En particular, se pasarán por alto ciertos detalles que se necesitan para la aplicación completa del algoritmo —por ejemplo, cómo encontrar una solución factible inicial de prueba—, que no son esenciales para la comprensión de los conceptos básicos. Las ideas se pueden resumir de la siguiente manera:

Concepto 1: Obtener, del *interior* de la región factible, una solución factible que lleve a la solución óptima.
Concepto 2: Moverse en la dirección que mejore el valor de la función objetivo lo más rápido posible.
Concepto 3: Transformar la región factible para colocar la solución de prueba actual cerca del centro, lo que permite una mejora grande cuando se aplique el concepto 2.

Para ilustrar estas ideas en esta sección se usará el siguiente ejemplo:

Maximizar $Z = x_1 + 2x_2,$

sujeta a

$$x_1 + x_2 \leq 8$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

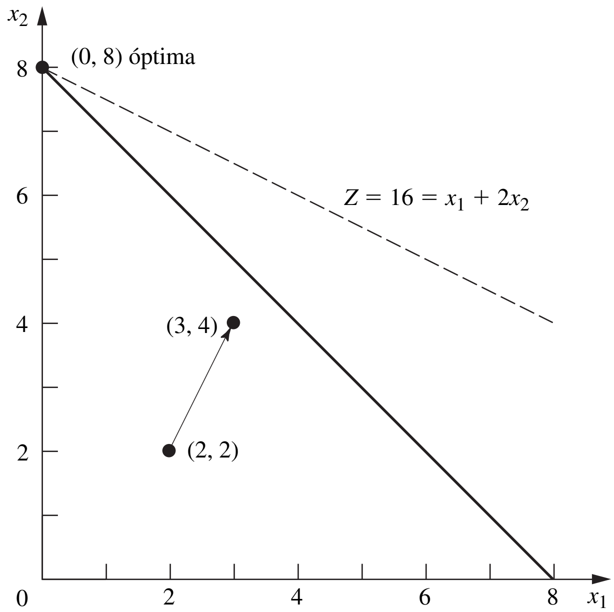
En la figura 7.3 se presenta la gráfica de este problema, donde se puede observar que la solución óptima es $(x_1, x_2) = (0, 8)$ con $Z = 16$. (Se describirá el significado de la flecha en la figura poco después.)

Se puede observar que el algoritmo de punto interior requiere una cantidad considerable de trabajo para resolver este ejemplo tan pequeño. La razón es que el algoritmo está diseñado para resolver problemas *muy grandes* de manera eficiente, pero es mucho menos eficiente que el método símplex (o el método gráfico en este caso) cuando se trata de resolver problemas pequeños. Sin embargo, mediante un ejemplo de sólo dos variables será posible mostrar de manera gráfica lo que hace el algoritmo.

Relevancia del gradiente para los conceptos 1 y 2

El algoritmo comienza con una solución de prueba inicial que —al igual que todas las soluciones de prueba subsecuentes— se encuentra en el *interior* de la región factible, es decir, *dentro de la frontera* de la región factible. Así, por ejemplo, la solución no debe estar en ninguna de las tres rectas ($x_1 = 0$, $x_2 = 0$, $x_1 + x_2 = 8$) que forman la frontera de esta región en la figura 7.3. (No se puede usar una solución de prueba que esté sobre la frontera porque ello conduciría a una operación

[6] En realidad, el enfoque básico de esta variante fue propuesto en 1967 por un matemático ruso, I. I. Dikin, y redescubierto, después de la publicación del trabajo de Karmarkar, por varios investigadores, entre los que se cuentan E. R. Barnes, T. M. Cavalier y A. L. Soyster. Se puede consultar además R. J. Vanderbei, M. S. Meketon y B. A. Freedman: "A Modification of Karmarkar's Linear Programming Algorithm", en *Algorithmica*, **1**(4) (Special Issue on New Approaches to Linear Programming): 395-407, 1986.



■ **FIGURA 7.3**
Ejemplo de algoritmo de punto interior.

matemática no definida de división entre cero en algún punto del algoritmo.) Se eligió $(x_1, x_2) = (2, 2)$ de manera arbitraria como la solución de prueba inicial.

Para comenzar con la aplicación de los conceptos 1 y 2, observe en la figura 7.3 que la dirección del movimiento desde el punto (2, 2) que aumenta el valor de Z a la mayor tasa posible es *perpendicular* (y hacia) la recta de la función objetivo, $Z = 16 = x_1 + 2x_2$. Esta dirección se indica con la flecha de (2, 2) a (3, 4). Si se suman los vectores se tiene

$$(3, 4) = (2, 2) + (1, 2),$$

donde el vector (1, 2) es el **gradiente** de la función objetivo. (En la sección 12.5 se estudiarán los gradientes con más detalle, dentro del contexto de *programación no lineal*, donde desde hace mucho se aplican algoritmos similares al de Karmarkar.) Los componentes de (1, 2) son exactamente los coeficientes de la función objetivo. De esta forma, con una modificación más, el gradiente (1, 2) define la dirección ideal para moverse; la pregunta sobre la distancia que *debe recorrerse* se considerará más adelante.

En realidad, el algoritmo opera sobre los problemas de programación lineal una vez que se han escrito en la forma aumentada. Sea x_3 la variable de holgura de la restricción funcional del ejemplo, esta forma es

$$\text{Maximizar} \qquad Z = x_1 + 2x_2,$$

sujeta a

$$x_1 + x_2 + x_3 = 8$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0.$$

En notación matricial —un poco diferente a la del capítulo 5, debido a que la variable de holgura está incorporada a la notación—, la forma aumentada se puede escribir en general como

$$\text{Maximizar} \qquad Z = \mathbf{c}^T\mathbf{x},$$

sujeta a

$$\mathbf{Ax} = \mathbf{b}$$

y

$$\mathbf{x} \geq \mathbf{0},$$

donde

$$\mathbf{c} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{A} = [1, \quad 1, \quad 1], \quad \mathbf{b} = [8], \quad \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

para este ejemplo. Observe que $\mathbf{c}^T = [1, 2, 0]$ es ahora el gradiente de la función objetivo.

En la figura 7.4 se muestra la gráfica de la forma aumentada del ejemplo. Ahora la región factible consiste en el triángulo con vértices (8, 0, 0), (0, 8, 0) y (0, 0, 8). Los puntos en el interior de esta región factible son aquellos en donde $x_1 > 0$, $x_2 > 0$ y $x_3 > 0$. Cada una de estas tres condiciones $x_j > 0$ tiene el efecto de forzar a (x_1, x_2) a alejarse de una de las tres rectas que forman la frontera de la región factible en la figura 7.3.

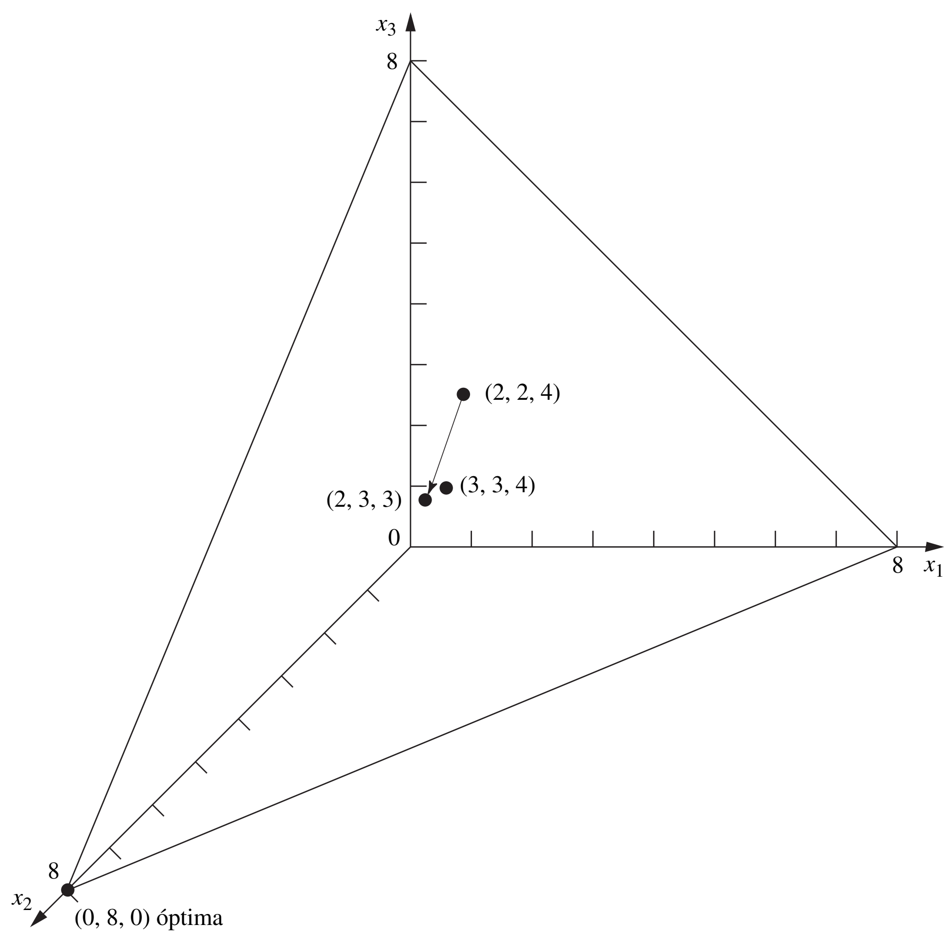
Utilización del gradiente proyectado para aplicar los conceptos 1 y 2

En la forma aumentada, la solución prueba inicial para el ejemplo es $(x_1, x_2, x_3) = (2, 2, 4)$. Si se agrega el gradiente (1, 2, 0) se obtiene

$$(3, 4, 4) = (2, 2, 4) + (1, 2, 0).$$

Sin embargo, ahora surgió una complicación. El algoritmo no se puede mover de (2, 2, 4) a (3, 4, 4) porque este último punto ¡es no factible! Si $x_1 = 3$ y $x_2 = 4$, entonces $x_3 = 8 - x_1 - x_2 = 1$ en lugar de 4. El punto (3, 4, 4) se encuentra en la arista de enfrente al ver el triángulo factible de la figura 7.4. Entonces, para que continúe factible, el algoritmo *proyecta* (de manera indirecta) el punto

FIGURA 7.4
Ejemplo en la forma aumentada de algoritmo de punto interior.



(3, 4, 4) al triángulo factible con el trazo de una recta *perpendicular* a este triángulo. Un vector de (0, 0, 0) a (1, 1, 1) es perpendicular a este triángulo, por lo que la recta perpendicular que pasa por (3, 4, 4) está dada por la ecuación

$$(x_1, x_2, x_3) = (3, 4, 4) - \theta(1, 1, 1),$$

donde θ es un escalar. Como el triángulo satisface la ecuación $x_1 + x_2 + x_3 = 8$, esta recta perpendicular intercepta al triángulo en (2, 3, 3). Como

$$(2, 3, 3) = (2, 2, 4) + (0, 1, -1),$$

el gradiente proyectado de la función objetivo —el **gradiente proyectado** sobre la región factible— es $(0, 1, -1)$. Este gradiente proyectado es el que define la dirección del movimiento del algoritmo (2, 2, 4), como lo indica la flecha en la figura 7.4.

Existe una fórmula disponible para calcular directamente el gradiente proyectado. Sea **P** la *matriz de proyección*

$$\mathbf{P} = \mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A},$$

el *gradiente proyectado* (en forma de columna) es

$$\mathbf{c}_p = \mathbf{P}\mathbf{c}.$$

Así, para el ejemplo,

$$\begin{aligned}\mathbf{P} &= \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix} - \begin{bmatrix}1\\1\\1\end{bmatrix}\left([1 \quad 1 \quad 1]\begin{bmatrix}1\\1\\1\end{bmatrix}\right)^{-1}[1 \quad 1 \quad 1]\\
&= \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix} - \frac{1}{3}\begin{bmatrix}1\\1\\1\end{bmatrix}[1 \quad 1 \quad 1]\\
&= \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix} - \frac{1}{3}\begin{bmatrix}1 & 1 & 1\\1 & 1 & 1\\1 & 1 & 1\end{bmatrix} = \begin{bmatrix}\frac{2}{3} & -\frac{1}{3} & -\frac{1}{3}\\-\frac{1}{3} & \frac{2}{3} & -\frac{1}{3}\\-\frac{1}{3} & -\frac{1}{3} & \frac{2}{3}\end{bmatrix},\end{aligned}$$

entonces,

$$\mathbf{c}_p = \begin{bmatrix}\frac{2}{3} & -\frac{1}{3} & -\frac{1}{3}\\-\frac{1}{3} & \frac{2}{3} & -\frac{1}{3}\\-\frac{1}{3} & -\frac{1}{3} & \frac{2}{3}\end{bmatrix}\begin{bmatrix}1\\2\\0\end{bmatrix} = \begin{bmatrix}0\\1\\-1\end{bmatrix}.$$

El movimiento de (2, 2, 4) en la dirección del gradiente proyectado $(0, 1, -1)$ implica un incremento de α a partir de cero en la fórmula

$$\mathbf{x} = \begin{bmatrix}2\\2\\4\end{bmatrix} + 4\alpha\mathbf{c}_p = \begin{bmatrix}2\\2\\4\end{bmatrix} + 4\alpha\begin{bmatrix}0\\1\\-1\end{bmatrix},$$

donde el coeficiente 4 se usa sólo para dar una cota superior de 1 a α para mantener la factibilidad (toda $x_j \geq 0$). Observe que al aumentar α a $\alpha = 1$ el resultado es que x_3 disminuye a $x_3 = 4 + 4(1)(-1) = 0$, donde $\alpha > 1$ conduce a $x_3 < 0$. Así, α mide la fracción usada de la distancia que se puede mover antes de dejar la región factible.

¿Cuánto debe crecer α para moverse a la siguiente solución de prueba? Como el incremento de Z es proporcional al de α, un valor cercano a 1 es bastante bueno para dar un paso relativamente grande hacia la optimalidad en la iteración actual. Sin embargo, el problema de un valor muy cercano a 1 es que la siguiente solución de prueba puede quedar acorralada con una frontera de restricción, lo que hace difícil introducir mejoras grandes en las iteraciones subsecuentes. Por tanto, es de gran ayuda que las soluciones de prueba queden cerca del centro de la región factible —o al

menos cerca del centro de la porción de la región factible en la vecindad de la solución óptima— y no demasiado cerca de ninguna frontera de restricción. Con esto en mente, Karmarkar estableció para su algoritmo que un valor de $\alpha = 0.25$ debe ser "seguro". En la práctica, a veces se usan valores mucho mayores (como $\alpha = 0.9$). Para desarrollar este ejemplo —y los problemas del final del capítulo— se eligió $\alpha = 0.5$. (En el IOR Tutorial se utiliza $\alpha = 0.5$ como valor dado, pero también se dispone de $\alpha = 0.9$.)

Esquema de centrado para implantar el concepto 3

Sólo falta un paso para completar la descripción del algoritmo: un esquema especial para transformar la región factible de manera que la solución de prueba actual quede cerca del centro. Se acaba de describir el beneficio de tener la solución prueba cerca del centro, pero otro beneficio importante del esquema de centrado es el cambio constante de la dirección del gradiente proyectado para que apunte más de cerca hacia una solución óptima, a medida que el algoritmo converge hacia esta solución.

Esta idea básica del esquema de centrado es directa, pues sólo se cambia la escala (unidades) de cada variable de manera que la solución de prueba sea equidistante de las fronteras de restricción en el nuevo sistema de coordenadas (el algoritmo original de Karmarkar utiliza un esquema de centrado más complejo).

En el ejemplo existen tres fronteras de restricción en la figura 7.3, cada una de las cuales corresponde a un valor de cero para cada una de las tres variables del problema en la forma aumentada, es decir, $x_1 = 0$, $x_2 = 0$ y $x_3 = 0$. En la figura 7.4 se puede observar que estas tres restricciones cortan el plano $\mathbf{Ax} = \mathbf{b}$ ($x_1 + x_2 + x_3 = 8$) para formar la frontera de la región factible. La solución de prueba inicial es $(x_1, x_2, x_3) = (2, 2, 4)$, y se encuentra dos unidades alejada de las restricciones $x_1 = 0$ y $x_2 = 0$ y alejada 4 unidades de $x_3 = 0$ si se usan las unidades de las variables respectivas. Sin embargo, cualesquiera que sean estas unidades, son arbitrarias y se pueden cambiar como se desee sin modificar el problema. Entonces, se dará la siguiente escala a las variables:

$$\tilde{x}_1 = \frac{x_1}{2}, \qquad \tilde{x}_2 = \frac{x_2}{2}, \qquad \tilde{x}_3 = \frac{x_3}{4}$$

con el fin de hacer que la solución prueba actual $(x_1, x_2, x_3) = (2, 2, 4)$ se convierta en

$$(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) = (1, 1, 1).$$

Con estas nuevas coordenadas —si se sustituye $2\tilde{x}_1$ por x_1, $2\tilde{x}_2$ por x_2, y $4\tilde{x}_3$ por x_3—, el problema se convierte en

$$\text{Maximizar} \qquad Z = 2\tilde{x}_1 + 4\tilde{x}_2,$$

sujeta a

$$2\tilde{x}_1 + 2\tilde{x}_2 + 4\tilde{x}_3 = 8$$

y

$$\tilde{x}_1 \geq 0, \qquad \tilde{x}_2 \geq 0, \qquad \tilde{x}_3 \geq 0,$$

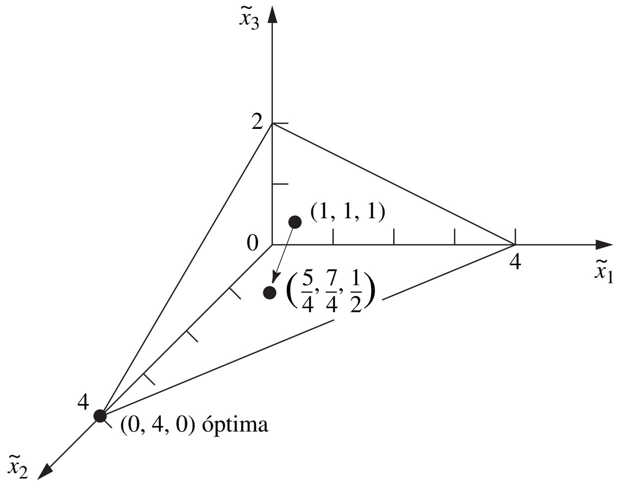
como se muestra de manera gráfica en la figura 7.5.

Observe que la solución prueba (1, 1, 1) de la figura 7.5 equidista de las fronteras de restricción $\tilde{x}_1 = 0, \tilde{x}_2 = 0, \tilde{x}_3 = 0$. En cada iteración subsiguiente se da una nueva escala al problema para que la solución de prueba sea siempre (1, 1, 1) en las coordenadas actuales.

Resumen e ilustración del algoritmo

A continuación se presenta un resumen y un ejemplo del algoritmo con la primera iteración del ejemplo, después con un resumen del procedimiento general y, por último, con la aplicación de este resumen a una segunda iteración.

Iteración 1. Dada la solución prueba inicial $(x_1, x_2, x_3) = (2, 2, 4)$, sea $\mathbf{D}$ la matriz diagonal correspondiente tal que $\mathbf{x} = \mathbf{D}\tilde{\mathbf{x}}$, de manera que



■ **FIGURA 7.5**
Ejemplo con la nueva escala de la iteración 1.

$$\mathbf{D} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

Entonces, las variables con la nueva escala son las componentes de

$$\tilde{\mathbf{x}} = \mathbf{D}^{-1}\mathbf{x} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ & 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{x_1}{2} \\ \frac{x_2}{2} \\ \frac{x_3}{4} \end{bmatrix}.$$

Con estas nuevas coordenadas, **A** y **c** se convierten en

$$\tilde{\mathbf{A}} = \mathbf{AD} = [1 \quad 1 \quad 1] \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix} = [2 \quad 2 \quad 4],$$

$$\tilde{\mathbf{c}} = \mathbf{Dc} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix}.$$

Entonces, la matriz de proyección es

$$\begin{aligned} \mathbf{P} &= \mathbf{I} - \tilde{\mathbf{A}}^T(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T)^{-1}\tilde{\mathbf{A}} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} \left([2 \quad 2 \quad 4] \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} \right)^{-1} [2 \quad 2 \quad 4] \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{24} \begin{bmatrix} 4 & 4 & 8 \\ 4 & 4 & 8 \\ 8 & 8 & 16 \end{bmatrix} = \begin{bmatrix} \frac{5}{6} & -\frac{1}{6} & -\frac{1}{3} \\ -\frac{1}{6} & \frac{5}{6} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}, \end{aligned}$$

de esta forma, el gradiente proyectado es

$$\mathbf{c}_p = \mathbf{P}\tilde{\mathbf{c}} = \begin{bmatrix} \frac{5}{6} & -\frac{1}{6} & -\frac{1}{3} \\ -\frac{1}{6} & \frac{5}{6} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}.$$

Sea v el *valor absoluto* de la componente *negativa* de $\mathbf{c}_p$ que tiene el *mayor* valor absoluto, entonces, $v = |-2| = 2$ en este caso. En consecuencia, en las coordenadas actuales, el algoritmo se mueve ahora de la solución de prueba actual $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) = (1, 1, 1)$ a la siguiente solución prueba

$$\tilde{\mathbf{x}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{\alpha}{v}\mathbf{c}_p = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{0.5}{2}\begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix} = \begin{bmatrix} \frac{5}{4} \\ \frac{7}{4} \\ \frac{1}{2} \end{bmatrix},$$

como se muestra en la figura 7.5. (Se escogió esta definición de v para hacer que la menor componente de $\tilde{\mathbf{x}}$ sea igual a cero cuando $\alpha = 1$ en esta ecuación para la siguiente solución prueba.) En las coordenadas originales, la solución es

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{D}\tilde{\mathbf{x}} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} \frac{5}{4} \\ \frac{7}{4} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{5}{2} \\ \frac{7}{2} \\ 2 \end{bmatrix}.$$

lo cual completa la iteración. Esta nueva solución se usará para comenzar con la siguiente iteración.

Los pasos que expusimos se pueden resumir como sigue para cualquier iteración.

Resumen del algoritmo de punto interior

1. Dada la solución prueba actual $(x_1, x_2, \ldots, x_n)$, se hace

$$\mathbf{D} = \begin{bmatrix} x_1 & 0 & 0 & \cdots & 0 \\ 0 & x_2 & 0 & \cdots & 0 \\ 0 & 0 & x_3 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & x_n \end{bmatrix}$$

2. Se calculan $\tilde{\mathbf{A}} = \mathbf{AD}$ y $\tilde{\mathbf{c}} = \mathbf{Dc}$.

3. Se calculan $\mathbf{P} = \mathbf{I} - \tilde{\mathbf{A}}^T(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T)^{-1}\tilde{\mathbf{A}}$ y $\mathbf{c}_p = \mathbf{P}\tilde{\mathbf{c}}$.

4. Se identifica la componente negativa de $\mathbf{c}_p$, que tiene el mayor valor absoluto y se hace igual a este valor. Después se calcula

$$\tilde{\mathbf{x}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \frac{\alpha}{v}\mathbf{c}_p,$$

donde α es una constante elegida entre 0 y 1 (por ejemplo, $\alpha = 0.5$).

5. Se calcula $\mathbf{x} = \mathbf{D}\tilde{\mathbf{x}}$ como la solución de prueba para la siguiente iteración (paso 1). (Si esta solución de prueba casi no cambia respecto de la anterior, se dice que el algoritmo converge a una solución óptima y se detiene.)

Ahora se aplicará este resumen a la iteración 2 del ejemplo.

Iteración 2

Paso 1:
Dada la solución actual $(x_1, x_2, x_3) = (\frac{5}{2}, \frac{7}{2}, 2)$, se hace

$$\mathbf{D} = \begin{bmatrix} \frac{5}{2} & 0 & 0 \\ 0 & \frac{7}{2} & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

(Observe que las variables con la nueva escala son

$$\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix} = \mathbf{D}^{-1}\mathbf{x} = \begin{bmatrix} \frac{2}{5} & 0 & 0 \\ 0 & \frac{2}{7} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{2}{5}x_1 \\ \frac{2}{7}x_2 \\ \frac{1}{2}x_3 \end{bmatrix},$$

de manera que las soluciones BF de estas nuevas coordenadas son

$$\tilde{\mathbf{x}} = \mathbf{D}^{-1}\begin{bmatrix} 8 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{16}{5} \\ 0 \\ 0 \end{bmatrix}, \qquad \tilde{\mathbf{x}} = \mathbf{D}^{-1}\begin{bmatrix} 0 \\ 8 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{16}{7} \\ 0 \end{bmatrix},$$

y

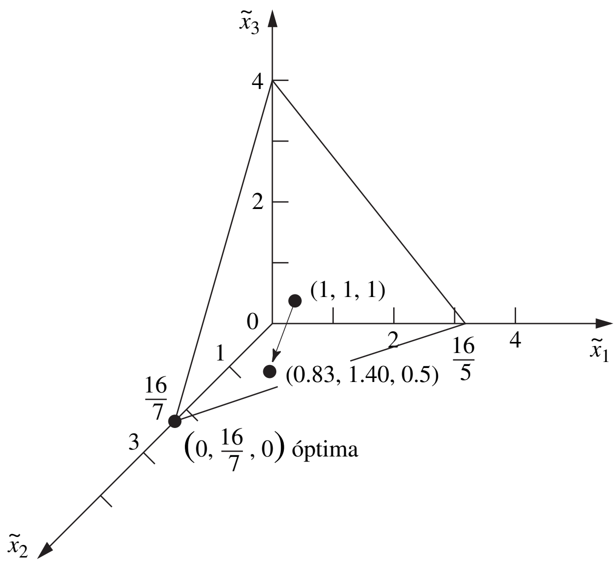
$$\tilde{\mathbf{x}} = \mathbf{D}^{-1}\begin{bmatrix} 0 \\ 0 \\ 8 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix},$$

como se muestra en la figura 7.6.)

Paso 2:

$$\tilde{\mathbf{A}} = \mathbf{AD} = [\tfrac{5}{2}, \tfrac{7}{2}, 2] \qquad \text{y} \qquad \tilde{\mathbf{c}} = \mathbf{Dc} = \begin{bmatrix} \frac{5}{2} \\ 7 \\ 0 \end{bmatrix}.$$

FIGURA 7.6
Ejemplo con la nueva escala de la iteración 2.



Paso 3:

$$\mathbf{P} = \begin{bmatrix} \frac{13}{18} & -\frac{7}{18} & -\frac{2}{9} \\ -\frac{7}{18} & \frac{41}{90} & -\frac{14}{45} \\ -\frac{2}{9} & -\frac{14}{45} & \frac{37}{45} \end{bmatrix} \quad \text{y} \quad \mathbf{c}_p = \begin{bmatrix} -\frac{11}{12} \\ \frac{133}{60} \\ -\frac{41}{15} \end{bmatrix}.$$

Paso 4:

$$\left|-\tfrac{41}{15}\right| > \left|-\tfrac{11}{12}\right|, \text{ entonces } v = \tfrac{41}{15} \text{ y}$$

$$\tilde{\mathbf{x}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{0.5}{\frac{41}{15}} \begin{bmatrix} -\frac{11}{12} \\ \frac{133}{60} \\ -\frac{41}{15} \end{bmatrix} = \begin{bmatrix} \frac{273}{328} \\ \frac{461}{328} \\ \frac{1}{2} \end{bmatrix} \approx \begin{bmatrix} 0.83 \\ 1.40 \\ 0.50 \end{bmatrix}.$$

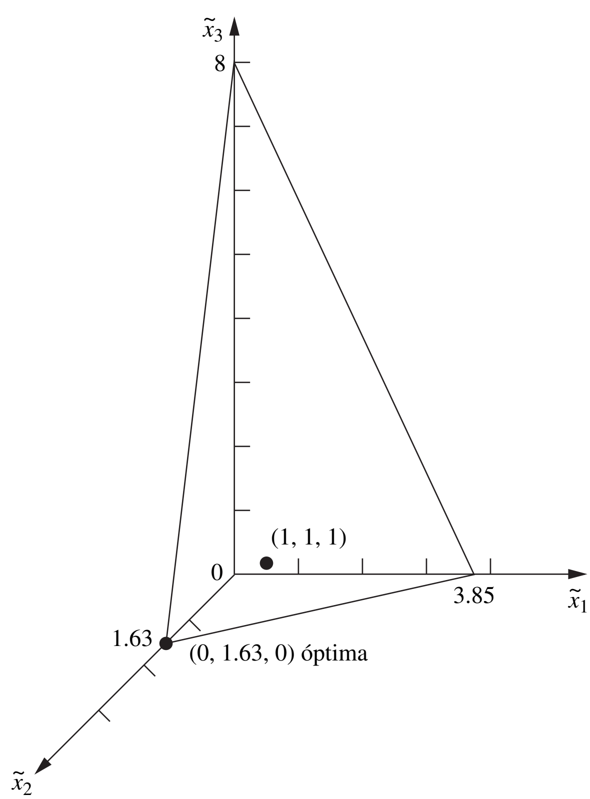
Paso 5:

$$\mathbf{x} = \mathbf{D}\tilde{\mathbf{x}} = \begin{bmatrix} \frac{1365}{656} \\ \frac{3227}{656} \\ 1 \end{bmatrix} \approx \begin{bmatrix} 2.08 \\ 4.92 \\ 1.00 \end{bmatrix}$$

es la solución de prueba de la iteración 3.

Como hay muy poco que aprender con la repetición de estos cálculos para otras iteraciones, no se harán más; pero en la figura 7.7 se presenta la región factible reconfigurada después de dar la nueva escala sobre la solución prueba que se acaba de obtener en la iteración 3. Como siempre, esta nueva escala coloca a la solución prueba en $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) = (1, 1, 1)$, equidistante de las fronteras de restricción $\tilde{x}_1 = 0, \tilde{x}_2 = 0$ y $\tilde{x}_3 = 0$. Observe en las figuras 7.5, 7.6 y 7.7 que la serie de iteraciones y las nuevas escalas tienen el efecto de "deslizar" la solución óptima hacia (1, 1, 1), mientras que

■ **FIGURA 7.7**
Ejemplo con la nueva escala de la iteración 3.



las otras soluciones BF tienden a alejarse. En algún momento, después de las iteraciones suficientes, la solución óptima quedará muy cerca de $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) = (0, 1, 0)$ después de dar la nueva escala, mientras que las otras soluciones BF estarán *muy* lejos del origen sobre los ejes $\tilde{x}_1$ y $\tilde{x}_3$. El paso 5 de esa iteración conducirá a una solución en las coordenadas originales muy cerca de la solución óptima $(x_1, x_2, x_3) = (0, 8, 0)$.

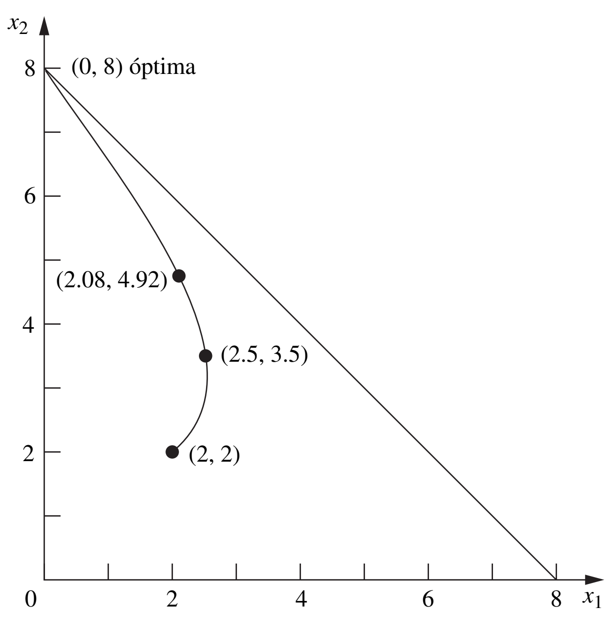
En la figura 7.8 se muestra el progreso del algoritmo en el sistema de coordenadas $x_1 = x_2$ original antes de aumentar el problema. Los tres puntos —$(x_1, x_2) = (2, 2)$, (2.5, 3.5) y (2.08, 4.92)— son las soluciones prueba para iniciar las respectivas iteraciones 1, 2 y 3. Se dibujó una curva suave a través de estos tres puntos y se continuó para mostrar la trayectoria del algoritmo en iteraciones subsecuentes a medida que se acerca a $(x_1, x_2) = (0, 8)$.

La restricción funcional de este ejemplo en particular es una desigualdad, pero las restricciones en forma de igualdad no causan problema al algoritmo, puesto que maneja las restricciones sólo después de ponerlas en la forma aumentada para convertirlas en igualdades ($\mathbf{Ax} = \mathbf{b}$). Para ilustrar este procedimiento, suponga que el único cambio en el ejemplo es que la restricción $x_1 + x_2 \leq 8$ se cambia a $x_1 + x_2 = 8$. Entonces, la región factible de la figura 7.3 cambia y queda sólo como el segmento de recta entre (8, 0) y (0, 8). Dada cualquier solución de prueba inicial en el interior ($x_1 > 0$ y $x_2 > 0$) de este segmento de recta —por ejemplo $(x_1, x_2) = (4, 4)$—, el algoritmo puede realizar los mismos cinco pasos dados en el resumen nada más con las dos variables y $\mathbf{A} = [1, 1]$. En cada iteración, el gradiente proyectado señala, a lo largo de este segmento, en la dirección de (0, 8). Con $\alpha = \frac{1}{2}$, la iteración 1 lleva de (4, 4) a (2, 6), la iteración 2 lleva de (2, 6) a (1, 7), etc. (En el problema 7.4-3 se pide verificar estos resultados.)

Aunque las dos versiones de este ejemplo tienen nada más una restricción, tener más implica sólo un cambio en el procedimiento como ya se vio (además del aumento de cálculos). Si se tiene una sola restricción, significa que $\mathbf{A}$ tiene un solo renglón, de manera que el término $(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T)^{-1}$ en el paso 3 se obtiene al tomar el recíproco del número obtenido del vector producto $\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T$. Si se tienen restricciones funcionales múltiples, $\mathbf{A}$ tiene varios renglones y entonces el término $(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T)^{-1}$ significa la inversa de la matriz que se obtuvo con el producto $\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T$.

Para concluir, es necesario hacer un comentario que puede dar una mejor perspectiva del algoritmo. En el caso de un ejemplo tan pequeño como el que se presentó, el algoritmo requiere un número relativamente grande de cálculos y después de muchas iteraciones obtiene sólo una aproximación de la solución óptima. Por el contrario, el procedimiento gráfico de la sección 3.1 encuentra de inmediato la solución óptima de la figura 7.3 y el método símplex requiere sólo una iteración rápida. Sin embargo, esto no debe ser razón para menospreciar la eficiencia del algoritmo de punto interior. Este algoritmo está diseñado para manejar problemas *grandes* que tienen muchos

FIGURA 7.8 Trayectoria del algoritmo de punto interior del ejemplo en el sistema de coordenadas original x_1-x_2.



cientos o miles de restricciones funcionales. El método símplex requiere miles de iteraciones para resolver este tipo de problemas. Cuando "obtiene una solución" en el interior de la región factible, el algoritmo de punto interior tiende a requerir un número mucho menor de iteraciones (aunque con mucho más trabajo por iteración). Por tanto, los algoritmos de punto interior similares al que se presentó tendrán un papel importante en el futuro de la programación lineal.

Vea en la sección 4.9 un análisis más detallado de este papel y una comparación del enfoque de punto interior con el método símplex. En dicha sección también se explican los papeles complementarios de los enfoques símplex y de punto interior, inclusive la forma en que se pueden combinar en un algoritmo híbrido.

Por último, se debe hacer hincapié en que esta sección ha proporcionado sólo una introducción conceptual al enfoque del punto interior de la programación lineal, es decir, se ha descrito una variante elemental del revolucionario algoritmo de Karmarkar de 1984. A lo largo de muchos años después, varios de los investigadores han desarrollado muchos avances importantes en el enfoque del punto interior. El estudio de este tema avanzado está fuera del alcance de este libro. Sin embargo, el lector interesado puede encontrar muchos detalles en las referencias seleccionadas que se enumeran al final de este capítulo.

7.5 CONCLUSIONES

El *método símplex dual* y la *programación lineal paramétrica* son valiosos en especial para el análisis postóptimo, aunque también son útiles en otros contextos.

La *técnica de la cota superior* proporciona una forma simplificada del método símplex para manejar la situación común en que muchas o todas las variables tienen cotas superiores explícitas. Reduce en una gran medida el esfuerzo computacional en problemas grandes.

Los paquetes de programación matemática casi siempre incluyen estos tres procedimientos y su uso es amplio. Debido a que su estructura básica se apoya en el método símplex según se presentó en el capítulo 4, conservan una eficiencia computacional excepcional para manejar problemas grandes como los descritos en la sección 4.8.

Se han desarrollado otros algoritmos con propósitos especiales que aprovechan la estructura especial de ciertos tipos de problemas de programación lineal (como los que se verán en los capítulos 8 y 9). En la actualidad se lleva a cabo una intensa investigación en esta área.

El *algoritmo de punto interior de Karmarkar* marca un nuevo desarrollo en programación lineal. Este algoritmo y sus variantes constituyen un nuevo enfoque poderoso y promisorio para resolver con eficiencia algunos problemas muy grandes.

REFERENCIAS SELECCIONADAS

1. Hooker, J. N., "Karmarkar's Linear Programming Algorithm", en *Interfaces*, **16**: 75-90, julio-agosto de 1986.
2. Luenberger, D. y Y. Ye, *Linear and Nolinear Programming*, 3a. ed., Springer, Nueva York, 2008.
3. Marsten, R., R. Subramanian, M. Saltzman, I. Lustig y D. Shanno, "Interior-Point Methods for Linear Programming: Just Call Newton, Lagrange, and Fiacco and McCormick!", en *Interfaces*, **20**: 105-116, julio-agosto de 1990.
4. Vanderbei, R. J., "Affine-Scaling for Linear Programs with Free Variables", en *Mathematical Programming*, **43**: 31-44, 1989.
5. Vanderbei, R. J., *Linear Programming: Foundations and Extensions*, 3a. ed., Springer, Nueva York, 2008.
6. Ye, Y., *Interior-Point Algorithms: Theory and Analysis*, Wiley, Nueva York, 1997.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos del capítulo 7

Rutinas interactivas en IOR Tutorial:

Introducción o revisión de un modelo general de programación lineal (Enter or Revise a General Linear Programming Model)
Preparación para el método símplex (Set up for the Simplex Method), sólo interactivo (Interactive Only)
Solución interactiva por el método símplex (Solve Interactively by the Simplex Method)
Método gráfico interactivo (Interactive Graphical Method)

Rutinas automáticas en IOR Tutorial:

Solución automática por el método símplex (Solve Automatically by the Simplex Method)
Solución automática por el algoritmo de punto interior (Solve Automatically by the Interior-Point Algorithm)
Método gráfico y análisis de sensibilidad (Graphical Method and Sensitivity Analysis)

Complemento de Excel:

Premium Solver for Education

Cap. 7, "Otros algoritmos para programación lineal", archivos para resolver los ejemplos:

Archivos de Excel
Archivo LINGO/LINDO
Archivo MPL/CPLEX

Glosario del capítulo 7

Complemento de este capítulo:

Programación Lineal por metas y sus procedimientos de solución (incluye dos casos: Una cura para Cuba y Seguridad en el aeropuerto)

Vea la documentación del software en el apéndice 1.

PROBLEMAS

Los símbolos a la izquierda de algunos problemas (o de sus incisos) significan lo siguiente:

I: Se sugiere que se use alguno de los procedimientos anteriores de IOR Tutorial (la impresión registra su trabajo). Para programación lineal paramétrica sólo se aplica al caso en que $\theta = 0$, después de lo cual debe proceder en forma manual.
C: Use la computadora para resolver el problema, y utilice la rutina automática para el algoritmo de punto interior en el IOR Tutorial.

Un asterisco en el número del problema indica que al final del libro se da al menos una respuesta parcial.

7.1-1. Considere el siguiente problema.

Maximizar $Z = -x_1 - 2x_2,$

sujeta a

$$\begin{aligned} 2x_1 + x_2 &\leq 40 \\ x_2 &\geq 15 \\ -2x_1 + x_2 &\leq 10 \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

I ***a***) Resuelva este problema en forma gráfica.
b) Use el *método símplex dual* para resolver el problema.
c) Trace en una gráfica la trayectoria que toma el método símplex dual.

7.1-2.* Use el *método símplex dual* para resolver el siguiente problema.

Minimizar $Z = 5x_1 + 2x_2 + 4x_3,$

sujeta a

$$\begin{aligned} 3x_1 + x_2 + 2x_3 &\geq 4 \\ 6x_1 + 3x_2 + 5x_3 &\geq 10 \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

7.1-3. Utilice el *método símplex dual* para resolver el siguiente problema.

Minimizar $Z = 7x_1 + 2x_2 + 5x_3 + 4x_4,$

sujeta a

$$\begin{aligned} 2x_1 + 4x_2 + 7x_3 + x_4 &\geq 5 \\ 8x_1 + 4x_2 + 6x_3 + 4x_4 &\geq 8 \\ 3x_1 + 8x_2 + x_3 + 4x_4 &\geq 4 \end{aligned}$$

y

$$x_j \geq 0, \quad \text{para } j = 1, 2, 3, 4.$$

7.1-4. Considere el siguiente problema.

Maximizar $Z = 5x_1 + 10x_2$,

sujeta a

$$3x_1 + x_2 \le 40$$
$$x_1 + x_2 \le 20$$
$$5x_1 + 3x_2 \le 90$$

y

$$x_1 \ge 0, \quad x_2 \ge 0.$$

I ***a)*** Resuelva mediante el *método símplex original* (en forma tabular). Identifique la solución básica *complementaria* del problema dual en cada iteración.

b) Resuelva el *dual* de este problema por el *método símplex dual*. Compare la secuencia de soluciones básicas resultante con las soluciones básicas complementarias que se obtuvieron en el inciso *a*).

7.1-5. Considere el ejemplo de análisis de sensibilidad del caso 1 de la sección 6.7, en donde se modifica la tabla símplex inicial de la tabla 4.8, al cambiar b_2 de 12 a 24, lo cual modifica los elementos respectivos de la columna del lado derecho de la tabla símplex *final* a 54, 6, 12 y −2. Comience con esta tabla símplex final modificada y utilice el *método símplex dual* para obtener la nueva solución óptima que se muestra en la tabla 6.21. Muestre su trabajo.

7.1-6.* Considere el inciso *a*) del problema 6.7-2. Utilice el *método símplex dual* para obtener la nueva solución óptima en cada caso; comience con la tabla símplex final revisada.

7.2-1.* Considere el siguiente problema.

Maximizar $Z = 8x_1 + 24x_2$,

sujeta a

$$x_1 + 2x_2 \le 10$$
$$2x_1 + x_2 \le 10$$

y

$$x_1 \ge 0, \quad x_2 \ge 0.$$

Suponga que Z representa la ganancia y que es posible modificar la función objetivo mediante un cambio apropiado del personal clave entre las dos actividades. En particular, suponga que la ganancia unitaria de la actividad 1 se puede aumentar a más de 8 (hasta un máximo de 18) a costa de disminuir la ganancia unitaria de la actividad 2 a menos de 24 en el doble de la cantidad. Entonces, en realidad Z se puede representar como

$$Z(\theta) = (8 + \theta)x_1 + (24 - 2\theta)x_2,$$

en donde θ es también una variable de decisión tal que $0 \le \theta \le 10$.

I ***a)*** Resuelva gráficamente la forma original de este problema. Después amplíe este procedimiento para resolver la extensión paramétrica del problema, esto es, encuentre el valor óptimo de $Z(\theta)$ como una función de θ, para $0 \le \theta \le 10$.

I ***b)*** Encuentre la solución óptima de la forma original por el método símplex. Después aplique *programación lineal paramétrica* para encontrar la solución óptima y el valor óptimo de $Z(\theta)$ como función de θ, para $0 \le \theta \le 10$. Grafique $Z(\theta)$.

c) Determine el valor óptimo de θ. Después indique de qué manera se pudo haber identificado este valor óptimo, si se hubieran resuelto en forma directa dos problemas de programación lineal normales. (*Sugerencia:* Una función convexa alcanza su máximo en un punto terminal.)

I **7.2-2.** Utilice *programación lineal paramétrica* para encontrar la solución óptima del siguiente problema, como una función de θ, para $0 \le \theta \le 20$.

Maximizar $Z(\theta) = (20 + 4\theta)x_1 + (30 - 3\theta)x_2 + 5x_3$,

sujeta a

$$3x_1 + 3x_2 + x_3 \le 10$$
$$8x_1 + 6x_2 + 4x_3 \le 25$$
$$6x_1 + x_2 + x_3 \le 15$$

y

$$x_1 \ge 0, \quad x_2 \ge 0, \quad x_3 \ge 0.$$

I **7.2-3.** Considere el siguiente problema.

Maximizar $Z(\theta) = (10 - \theta)x_1 + (12 + \theta)x_2 + (7 + 2\theta)x_3$,

sujeta a

$$x_1 + 2x_2 + 2x_3 \le 30$$
$$x_1 + x_2 + x_3 \le 20$$

y

$$x_1 \ge 0, \quad x_2 \ge 0, \quad x_3 \ge 0.$$

a) Utilice *programación lineal paramétrica* para encontrar la solución óptima de este problema, como una función de θ, para $\theta \ge 0$.

b) Construya el modelo dual de este problema. Después, encuentre la solución óptima para ese problema dual como una función de θ, para $\theta \ge 0$, con el uso del método descrito en la última parte de la sección 7.2. Indique en una gráfica qué hace este procedimiento. Compare las soluciones obtenidas con las soluciones básicas complementarias que se obtuvieron en el inciso *a*).

I **7.2-4.*** Mediante el procedimiento de *programación lineal paramétrica* haga cambios sistemáticos en los parámetros b_i para encontrar una solución óptima del problema siguiente, como una función de θ, para $0 \le \theta \le 25$.

Maximizar $Z(\theta) = 2x_1 + x_2$,

sujeta a

$$x_1 \le 10 + 2\theta$$
$$x_1 + x_2 \le 25 - \theta$$
$$x_2 \le 10 + 2\theta$$

y

$$x_1 \ge 0, \quad x_2 \ge 0.$$

Indique en una gráfica qué hace este procedimiento.

I **7.2-5.** Utilice *programación lineal paramétrica* para encontrar una solución óptima del siguiente problema, como una función de θ, para $0 \le \theta \le 30$.

Maximizar $Z(\theta) = 5x_1 + 42x_2 + 28x_3 + 49x_4$,

sujeta a

$$3x_1 - 2x_2 + x_3 + 3x_4 \le 135 - 2\theta$$
$$2x_1 + 4x_2 - x_3 + 2x_4 \le 78 - \theta$$
$$x_1 + 2x_2 + x_3 + 2x_4 \le 30 + \theta$$

y

$$x_j \ge 0, \quad \text{para } j = 1, 2, 3, 4.$$

Después identifique el valor de θ que da el valor óptimo mayor de $Z(\theta)$.

7.2-6. Considere el problema 6.7-3. Utilice *programación lineal paramétrica* para encontrar la solución óptima como una función de θ para $-20 \leq \theta \leq 0$. (*Sugerencia:* Cambie $-\theta'$ por θ, y después incremente el valor de θ' a partir de cero.)

7.2-7. Considere la función $Z^*(\theta)$ que se muestra en la figura 7.1 para *programación lineal paramétrica* con cambios sistemáticos en los parámetros c_j.

a) Explique por qué esta función es lineal por partes.
b) Muestre que esta función debe ser convexa.

7.2-8. Considere la función $Z^*(\theta)$ que se muestra en la figura 7.2 para *programación lineal paramétrica* con cambios sistemáticos en los parámetros b_i.

a) Explique por qué esta función es lineal por partes.
b) Muestre por qué esta función debe ser cóncava.

7.2-9. Sea

$$Z^* = \text{máx} \left\{ \sum_{j=1}^{n} c_j x_j \right\},$$

sujeta a

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i, \quad \text{para } i = 1, 2, \ldots, m,$$

y

$$x_j \geq 0, \quad \text{para } j = 1, 2, \ldots, n$$

(donde las a_{ij}, b_i y c_j son constantes fijas) y sea $(y_1^*, y_2^*, \ldots, y_m^*)$ la solución óptima dual correspondiente. Sea

$$Z^{**} = \text{máx} \left\{ \sum_{j=1}^{n} c_j x_j \right\},$$

sujeta a

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i + k_i, \quad \text{para } i = 1, 2, \ldots, m,$$

y

$$x_j \geq 0, \quad \text{para } j = 1, 2, \ldots, n,$$

donde $k_1, k_2, \ldots, k_m$ son las constantes dadas. Demuestre que

$$Z^{**} \leq Z^* + \sum_{i=1}^{m} k_i y_i^*.$$

7.3-1. Considere el siguiente problema.

$$\text{Maximizar} \quad Z = 2x_1 + 3x_2,$$

sujeta a

$$\begin{aligned} 3x_1 - 9x_2 &\leq 20 \\ 3x_1 \qquad &\leq 40 \\ 9x_2 &\leq 40 \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

I ***a***) Resuelva este problema por el método gráfico.
b) Utilice la *técnica de la cota superior* para resolver este problema de forma manual.
c) Trace en una gráfica la trayectoria que sigue la técnica de la cota superior.

7.3-2.* Utilice la *técnica de la cota superior* para resolver el siguiente problema de forma manual.

$$\text{Maximizar} \quad Z = x_1 + 3x_2 - 2x_3,$$

sujeta a

$$\begin{aligned} x_2 - 2x_3 &\leq 1 \\ 2x_1 + x_2 + 2x_3 &\leq 8 \\ x_1 \qquad &\leq 1 \\ x_2 \qquad &\leq 3 \\ x_3 &\leq 2 \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

7.3-3. Utilice la *técnica de la cota superior* a mano para resolver el siguiente problema.

$$\text{Maximizar} \quad Z = 2x_1 + 3x_2 - 2x_3 + 5x_4,$$

sujeta a

$$\begin{aligned} 2x_1 + 2x_2 + x_3 + 2x_4 &\leq 5 \\ x_1 + 2x_2 - 3x_3 + 4x_4 &\leq 5 \end{aligned}$$

y

$$0 \leq x_j \leq 1, \quad \text{para } j = 1, 2, 3, 4.$$

7.3-4. Utilice la *técnica de la cota superior* para resolver el siguiente problema de forma manual.

$$\text{Maximizar} \quad Z = 2x_1 + 5x_2 + 3x_3 + 4x_4 + x_5,$$

sujeta a

$$\begin{aligned} x_1 + 3x_2 + 2x_3 + 3x_4 + x_5 &\leq 6 \\ 4x_1 + 6x_2 + 5x_3 + 7x_4 + x_5 &\leq 15 \end{aligned}$$

y

$$0 \leq x_j \leq 1, \quad \text{para } j = 1, 2, 3, 4, 5.$$

7.3-5. Aplique la *técnica de la cota superior* y el *método símplex dual* para resolver el siguiente problema de forma manual.

$$\text{Minimizar} \quad Z = 3x_1 + 4x_2 + 2x_3,$$

sujeta a

$$\begin{aligned} x_1 + x_2 + x_3 &\geq 15 \\ x_2 + x_3 &\geq 10 \end{aligned}$$

y

$$0 \leq x_1 \leq 25, \quad 0 \leq x_2 \leq 5, \quad 0 \leq x_3 \leq 15.$$

C **7.4-1.** Reconsidere el ejemplo que se utilizó para ilustrar el algoritmo de punto interior en la sección 7.4. Suponga que se usó $(x_1, x_2) = (1, 3)$ como solución de prueba factible inicial. Realice dos iteraciones a mano a partir de esta solución. Después use la rutina automática del IOR Tutorial para verificar su trabajo.

7.4-2. Considere el siguiente problema.

Maximizar $Z = 3x_1 + x_2$,

sujeta a

$x_1 + x_2 \leq 4$

y

$x_1 \geq 0, \quad x_2 \geq 0.$

I ***a***) Resuelva este problema en forma gráfica. Además, identifique todas las soluciones FEV.

C ***b***) Inicie con la solución de prueba inicial $(x_1, x_2) = (1, 1)$, para realizar a mano cuatro iteraciones del algoritmo de punto interior que se presentó en la sección 7.4. Después, use la rutina automática del IOR Tutorial para verificar su trabajo.

c) Dibuje figuras correspondientes a las figuras 7.4, 7.5, 7.6, 7.7 y 7.8 para representar este problema. En cada caso, identifique las soluciones básicas (o en un vértice) factibles en el sistema de coordenadas actual. (Se pueden usar las soluciones de prueba para determinar los gradientes proyectados.)

7.4-3. Considere el siguiente problema.

Maximizar $Z = x_1 + 2x_2$,

sujeta a

$x_1 + x_2 = 8$

y

$x_1 \geq 0, \quad x_2 \geq 0.$

C ***a***) Casi al final de la sección 7.4 se estudia qué hace el algoritmo de punto interior sobre este problema si comienza con la solución prueba inicial $(x_1, x_2) = (4, 4)$. Verifique los resultados que se obtienen mediante la realización de dos iteraciones de forma manual. Después use la rutina automática del IOR Tutorial para comprobar sus resultados.

b) Utilice estos resultados para predecir cuáles serían las soluciones de prueba subsecuentes si se realizaran más iteraciones.

c) Suponga que la regla de detención adoptada por el algoritmo en esta aplicación es que el algoritmo se detiene cuando dos soluciones de prueba sucesivas difieren en menos de 0.01 en cualquiera de sus componentes. Utilice las predicciones del inciso *b*) para pronosticar la solución de prueba final y el número total de iteraciones que se requieren para llegar a ella. ¿Qué tan cerca estaría esta solución de la solución óptima $(x_1, x_1) = (0, 8)$?

7.4-4. Considere el siguiente problema.

Maximizar $Z = 3x_1 + x_2$,

sujeta a

$3x_1 + 2x_2 \leq 45$
$6x_1 + x_2 \leq 45$

y

$x_1 \geq 0, \quad x_2 \geq 0.$

I ***a***) Resuelva este problema en forma gráfica.

b) Encuentre el gradiente de la función objetivo del sistema de coordenadas original x_1-x_2. Si se mueve del origen en la dirección del gradiente hasta alcanzar la frontera de la región factible, ¿a dónde se llega en relación con la solución óptima?

C ***c***) Si se inicia con la solución prueba inicial $(x_1, x_2) = (1, 1)$, use el IOR Tutorial para realizar 10 iteraciones del algoritmo de punto interior que se presentó en la sección 7.4.

C ***d***) Repita el inciso *c*) con $\alpha = 0.9$.

7.4-5. Considere el siguiente problema.

Maximizar $Z = 2x_1 + 5x_2 + 7x_3$,

sujeta a

$x_1 + 2x_2 + 3x_3 = 6$

y

$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$

I ***a***) Grafique la región factible.

b) Encuentre el *gradiente* de la función objetivo y el *gradiente proyectado* sobre la región factible.

c) Comience con la solución de prueba inicial $(x_1, x_2, x_3) = (1, 1, 1)$, y realice a mano dos iteraciones del algoritmo de punto interior que se presentó en la sección 7.4.

C ***d***) Comience con la misma solución de prueba inicial y use el IOR Tutorial para realizar 10 iteraciones.

C **7.4-6.** Comience con la solución prueba inicial $(x_1, x_2) = (2, 2)$, y utilice el IOR Tutorial para realizar 15 iteraciones del algoritmo de punto interior que se presentó en la sección 7.4 en el caso del problema de la Wyndor Glass Co. de la sección 3.1. También dibuje una figura como la 7.8 para mostrar la trayectoria del algoritmo en el sistema de coordenadas original x_1-x_2.

CAPÍTULO 8

Problemas de transporte y asignación

En el capítulo 3 se hizo hincapié en la amplia variedad de aplicaciones de la programación lineal. En éste se ampliará el horizonte con la presentación de dos tipos particularmente importantes (y relacionados) de problemas de programación lineal. El primero de ellos se conoce como *problema de transporte*, nombre que recibe porque muchas de sus aplicaciones involucran cómo determinar la manera óptima de transportar bienes. Sin embargo, algunas de sus aplicaciones importantes —como la programación de la producción— en realidad no tienen nada que ver con el transporte.

El segundo tipo, llamado *problema de asignación*, incluye aplicaciones como la asignación de personas a tareas. Aunque sus usos parecen diferir de los del problema de transporte, se verá que los asuntos de asignación se pueden considerar un caso especial del problema de transporte.

En el siguiente capítulo se introducirán otros tipos especiales de programación lineal que involucran *redes*, como el *problema del flujo de costo mínimo* (sección 9.6). Ahí se verá que, en realidad, los problemas de transporte y de asignación son casos especiales del problema de flujo de costo mínimo. En este capítulo se estudia la representación de redes de los problemas de transporte y asignación.

Las aplicaciones de los problemas de transporte y asignación tienden a requerir un número muy grande de restricciones y variables, de manera que una solución en computadora del método símplex puede necesitar de un esfuerzo computacional exorbitante. Por fortuna, una característica clave de estos problemas es que la mayor parte de los coeficientes a_{ij} de las restricciones son iguales a cero. Como resultado, se han podido desarrollar algoritmos *simplificados* especiales que logran ahorros computacionales sorprendentes para explotar esta estructura especial del problema. En consecuencia, es importante familiarizarse bien con estos tipos especiales de problemas a fin de reconocerlos cuando surjan y aplicar el procedimiento adecuado para resolverlos.

TABLA 8.1 Tabla de coeficientes de las restricciones de programación lineal

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Para describir las estructuras especiales se introducirá la tabla (o matriz) de coeficientes que se muestra en la tabla 8.1, en donde a_{ij} es el coeficiente de la j-ésima variable de la i-ésima restricción funcional. Más adelante, las partes de la tabla que contienen coeficientes iguales a cero se dejarán en blanco, mientras que los bloques en donde hay coeficientes distintos a cero se indicarán con áreas sombreadas.

Después de presentar un ejemplo prototipo del problema de transporte, se describirá la estructura especial de este modelo y se darán ejemplos adicionales de sus aplicaciones. En la sección 8.2 se presenta el *método símplex de transporte*, como una versión especial simplificada del método símplex para resolver estos problemas con eficiencia. (En la sección 9.7 se verá que este algoritmo está relacionado con el *método símplex de redes*, otra versión simplificada del método símplex para la solución eficiente de cualquier problema de flujo de costo mínimo, incluso problemas de transporte y de asignación.) La sección 8.3 está dedicada al problema de asignación. Después, en la sección 8.4 se presenta un algoritmo especializado, llamado *algoritmo húngaro*, para resolver de manera muy eficiente sólo problemas de asignación.

Recuadro de aplicación

Procter & Gamble (P & G) produce y comercializa más de 300 marcas de productos a nivel mundial. Esta compañía ha crecido continuamente a través de su larga historia que data desde 1830. Para mantener y acelerar ese crecimiento, un importante estudio de IO se realizó para fortalecer la efectividad global de P & G. Antes del estudio, la cadena de suministro de la compañía consistía en cientos de proveedores para las 50 categorías de productos en 60 plantas, 15 centros de distribución y más de 1 000 zonas de consumo. Sin embargo, en la medida en que la compañía consideró marcas globales, la administración se percató de que se requería consolidar las plantas a fin de reducir los gastos de manufactura, mejorar la velocidad de entrega al mercado y reducir la inversión de capital. Por lo cual, el estudio se enfocó al rediseño del sistema de producción y distribución de la compañía para sus operaciones en Norteamérica. El resultado fue una reducción del número de plantas en Norteamérica de casi 20 por ciento, *ahorrando más* de **200 millones de dólares** en costos antes de impuestos *por año*.

Gran parte del estudio consistió en la formulación y solución de problemas de transporte de categorías individuales de producto. Para cada opción referente a mantener abiertas ciertas plantas, la solución del correspondiente problema de transporte para cierta categoría de producto mostró cuál sería el costo de distribución para enviar dicha categoría de producto desde esas plantas hacia los centros de distribución y zonas de consumo.

Fuente: D. Camm, T. E. Chorman, F. A. Dill, J. R. Evans, D. J. Sweeney y G. W. Wegryn: "Blending OR/MS, Judgement, and GIS: Restructuring P & G Supply Chain", *Interfaces*, **27**(1): 128-142, enero-febrero, 1997. (En nuestro sitio de internet, www.mhhe.com/hillier, aparece una liga de este artículo.)

El sitio de internet de este libro también contiene un complemento de este capítulo. Es un estudio de caso completo —incluye el análisis— que ilustra la forma en la cual una decisión corporativa de dónde se debe localizar una planta —una refinería de petróleo en este caso— puede requerir la resolución de muchos problemas de transporte. (Uno de los casos de este capítulo pide que se continúe con el análisis de una extensión de este caso de estudio.)

8.1 PROBLEMA DE TRANSPORTE

Ejemplo prototípico

Uno de los productos más importantes de la P & T COMPANY es el chícharo enlatado. Los chícharos se preparan en tres enlatadoras —cercanas a Bellingham, Washington; Eugene, Oregon, y Albert Lea, Minnesota— y después se envían por camión a cuatro almacenes de distribución —Sacramento, California; Salt Lake City, Utah; Rapid City, South Dakota, y Albuquerque, Nuevo México— en el oeste de Estados Unidos, como se muestra en la figura 8.1. Debido a que los costos de embarque constituyen un gasto importante, la administración ha iniciado un estudio para reducirlos a su mínima expresión. Se ha estimado la producción de cada enlatadora durante la próxima temporada y se ha asignado a cada almacén cierta cantidad de la producción total de chícharos. En la tabla 8.2 se proporciona esta información —en unidades de carga de camión—, junto con el costo de transporte por camión cargado de cada combinación de enlatadora-almacén. Como se ve, hay un total de 300 cargas de camión que se deben transportar. El problema es determinar el plan de asignación de estos embarques a las distintas combinaciones de enlatadora-almacén que *minimice el costo total de transporte*.

Si se hace caso omiso de la distribución geográfica de las enlatadoras y los almacenes es posible proporcionar una sencilla *representación en red* del problema. Para ello, se deben alinear las enlatadoras en una columna a la izquierda y los almacenes en otra columna a la derecha. Esta representación se muestra en la figura 8.2. Las flechas indican las probables rutas de los camiones, donde el número junto a cada flecha es el costo de envío por carga de camión por esa ruta. Los números entre los paréntesis cuadrados junto a cada localidad son las cargas de camión que deben enviarse *desde ese lugar*, esto es, la asignación que debe llegar a cada almacén está dada como números negativos.

En realidad, el problema descrito en la figura 8.2 es de programación lineal del *tipo de los problemas de transporte*. Para formularlo, sea Z el costo total de transporte y sea x_{ij} ($i = 1, 2, 3;$



■ **FIGURA 8.1**
Localización de enlatadoras y almacenes del problema de la P & T Co.

$j = 1, 2, 3, 4$) el número de cargas de camión enviadas de la enlatadora i al almacén j. El objetivo es seleccionar valores de estas 12 variables de decisión (las x_{ij}) para

$$\text{Minimizar} \quad Z = 464x_{11} + 513x_{12} + 654x_{13} + 867x_{14} + 352x_{21} + 416x_{22} + 690x_{23} + 791x_{24} + 995x_{31} + 682x_{32} + 388x_{33} + 685x_{34},$$

sujeto a las restricciones

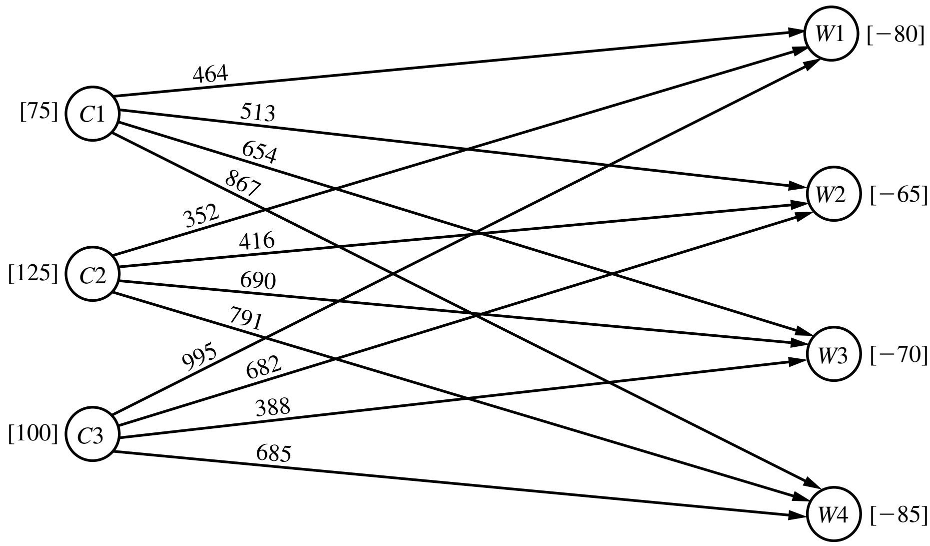
$$\begin{aligned}
&x_{11} + x_{12} + x_{13} + x_{14} &&= 75\\
&x_{21} + x_{22} + x_{23} + x_{24} &&= 125\\
&x_{31} + x_{32} + x_{33} + x_{34} &&= 100\\
&x_{11} + x_{21} + x_{31} &&= 80\\
&x_{12} + x_{22} + x_{32} &&= 65\\
&x_{13} + x_{23} + x_{33} &&= 70\\
&x_{14} + x_{24} + x_{34} &&= 85
\end{aligned}$$

y

$$x_{ij} \geq 0 \qquad (i = 1, 2, 3;\ j = 1, 2, 3, 4).$$

■ **TABLA 8.2** Datos de transporte de P & T Co.

	Costo de embarque ($) por carga				
	Almacén				
	1	**2**	**3**	**4**	**Producción**
Enlatadora 1	464	513	654	867	75
2	352	416	690	791	125
3	995	682	388	685	100
Asignación	80	65	70	85	



■ **FIGURA 8.2**
Representación de red del problema de la P & T Co.

En la tabla 8.3 se muestran los coeficientes de las restricciones. Como se verá después en esta sección, lo que caracteriza a la situación dada como un problema de transporte es la estructura especial del patrón de sus coeficientes, no su contexto. Sin embargo, primero se describirán varias características del modelo del problema de transporte.

Modelo del problema de transporte

Para describir el modelo general del problema de transporte es necesario emplear términos mucho menos específicos que los que se usaron para designar los componentes del ejemplo prototípico. En particular, el problema general de transporte se refiere —en sentido literal o figurado— a la distribución de *cualquier* mercancía desde *cualquier* grupo de centros de suministro, llamados **orígenes**, a *cualquier* grupo de centros de recepción, llamados **destinos**, de tal manera que se minimicen los costos totales de distribución. La correspondencia en terminología entre el ejemplo prototípico y el problema general se resume en la tabla 8.4.

Como lo indican el cuarto y quinto renglones de la tabla, cada origen tiene que distribuir cierto **suministro** de unidades a los destinos y cada destino tiene cierta **demanda** de unidades que deben recibir de los orígenes. El modelo de un problema de transporte se basa en los siguientes supuestos sobre estos suministros y demandas.

> **Supuesto de requerimientos:** Cada origen tiene un *suministro* fijo de unidades, el cual debe distribuirse completo entre los destinos. (Con s_i se denota el número de unidades que suministra el origen i, para $i = 1, 2, \ldots, m$.) De manera similar, el destino tiene una *demanda* fija de unidades, y debe satisfacerse por los orígenes. (Con d_j se denota el número de unidades que recibe el destino j, para $j = 1, 2, \ldots, n$.)

■ **TABLA 8.3** Coeficientes de restricción de P & T Co.

	Coeficiente de:												
	x_{11}	x_{12}	x_{13}	x_{14}	x_{21}	x_{22}	x_{23}	x_{24}	x_{31}	x_{32}	x_{33}	x_{34}	
	1	1	1	1									Restricciones de enlatadora
					1	1	1	1					
									1	1	1	1	
A =	1				1				1				Restricciones de almacén
		1				1				1			
			1				1				1		
				1				1				1	

TABLA 8.4 Terminología del problema de transporte

Ejemplo prototípico	Problema general
Cargas de camión de chícharos enlatados	Unidades de un bien
Tres enlatadoras	m orígenes
Cuatro almacenes	n destinos
Producción de la enlatadora i	s_i recursos en el origen i
Asignación al almacén j	Demanda d_j del destino j
Costo de envío por carga desde la enlatadora i hasta el almacén j	Costo c_{ij} por unidad distribuida desde el origen i al destino j

Este supuesto sirve para el problema de la P & T Co., ya que cada enlatadora (origen) tiene una producción fija y cada almacén (destino) tiene una asignación fija.

El supuesto de que no hay margen en las cantidades que deben enviarse o recibirse significa que es necesario un balance entre el suministro total de todos los orígenes y la demanda total de todos los destinos.

Propiedad de soluciones factibles: un problema de transporte tiene soluciones factibles si y sólo si

$$\sum_{i=1}^{m} s_i = \sum_{j=1}^{n} d_j.$$

Por fortuna, estas sumas son iguales en el caso de P & T Co., puesto que la tabla 8.2 indica que los suministros (producción) suman 300 camiones cargados, esto es, igual que el número de las demandas (asignaciones).

En algunos problemas reales, los suministros representan cantidades *máximas* —y no cantidades fijas— que deben distribuirse. En forma parecida, en otros casos, las demandas representan cantidades máximas —y no fijas— que deben recibirse. Tales problemas no se ajustan por completo al problema de transporte porque violan el *supuesto de requerimientos*. Sin embargo, es posible *reformular* el problema de manera que se ajuste al modelo con la introducción de un *destino ficticio* o un *origen ficticio* que considere la holgura entre las cantidades reales distribuidas y las máximas. Esta situación se ilustrará con dos ejemplos al final de esta sección.

El último renglón de la tabla 8.4 se refiere a un costo por unidad distribuida. Este *costo unitario* implica el siguiente supuesto básico de cualquier problema de transporte.

Supuesto de costo: El costo de distribuir unidades de un origen a un destino dados es *directamente proporcional* al número de unidades distribuidas. Por tanto, este costo es igual al *costo unitario* de distribución *multiplicado* por el *número de unidades distribuidas*. (El costo unitario del origen i al destino j se denota por c_{ij}.)

Este supuesto se cumple en el caso del problema de la P & T Co., en razón de que el costo de envío de los chícharos de una enlatadora a un centro de distribución es directamente proporcional al número de camiones embarcados.

Los únicos datos necesarios para elaborar un modelo de transporte son suministros, demandas y costos unitarios. Éstos son los *parámetros del modelo*. Todos estos parámetros se pueden resumir de manera conveniente en la *tabla de parámetros* que se muestra en el cuadro 8.5.

El modelo: Cualquier problema —ya sea que involucre el transporte o no— se ajusta a este modelo de un problema de transporte si se puede describir por completo en términos de una *tabla de parámetros* como la que se presenta en la tabla 8.5 y satisface tanto el *supuesto de requerimientos* como el *de costo*. El objetivo es minimizar el costo total de distribuir las unidades. Todos los parámetros del modelo están incluidos en esta tabla de parámetros.

En consecuencia, la formulación de cierto problema como uno de transporte sólo requiere llenar una tabla de parámetros en el formato de la tabla 8.5. (La tabla de parámetros del problema de la P & T Co., se muestra en la figura 8.2.) De manera alterna, la misma información se puede

TABLA 8.5 Tabla de parámetros del problema de transporte

		Costo por unidad distribuida				
		Destino				
		1	2	⋯	n	Recursos
Origen	1	c_{11}	c_{12}	⋯	c_{1n}	s_1
	2	c_{21}	c_{22}	⋯	c_{2n}	s_2
	⋮	⋯	⋯	⋯	⋯	⋮
	m	c_{m1}	c_{m2}	⋯	c_{mn}	s_m
Demanda		d_1	d_2	⋯	d_n	

proporcionar mediante la representación en red del problema que se muestra en la figura 8.3, como ya se hizo en la figura 8.2 en el problema de la P & T Co. Algunos problemas que no tienen relación con transporte también se pueden formular como un problema de este tipo en cualquiera de las dos formas mencionadas. En la sección Worked Examples en el sitio de internet de este libro se incluye **otro ejemplo** de este tipo de problemas.

Como un problema de transportación se puede formular sólo con llenar una tabla de parámetros o mediante su representación en red, no es necesario un modelo matemático formal. Sin embargo, se mostrará el modelo del problema de transporte general sólo una vez para hacer notar que es un tipo especial de problema de programación lineal.

Sea Z el costo total de distribución y x_{ij} ($i = 1, 2, \ldots, m; j = 1, 2, \ldots, n$) el número de unidades que se distribuyen del origen i al destino j, la formulación de programación lineal de este problema es

$$\text{Minimizar} \qquad Z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}x_{ij},$$

sujeta a

$$\sum_{j=1}^{n} x_{ij} = s_i \qquad \text{para } i = 1, 2, \ldots, m,$$

$$\sum_{i=1}^{m} x_{ij} = d_j \qquad \text{para } j = 1, 2, \ldots, n,$$

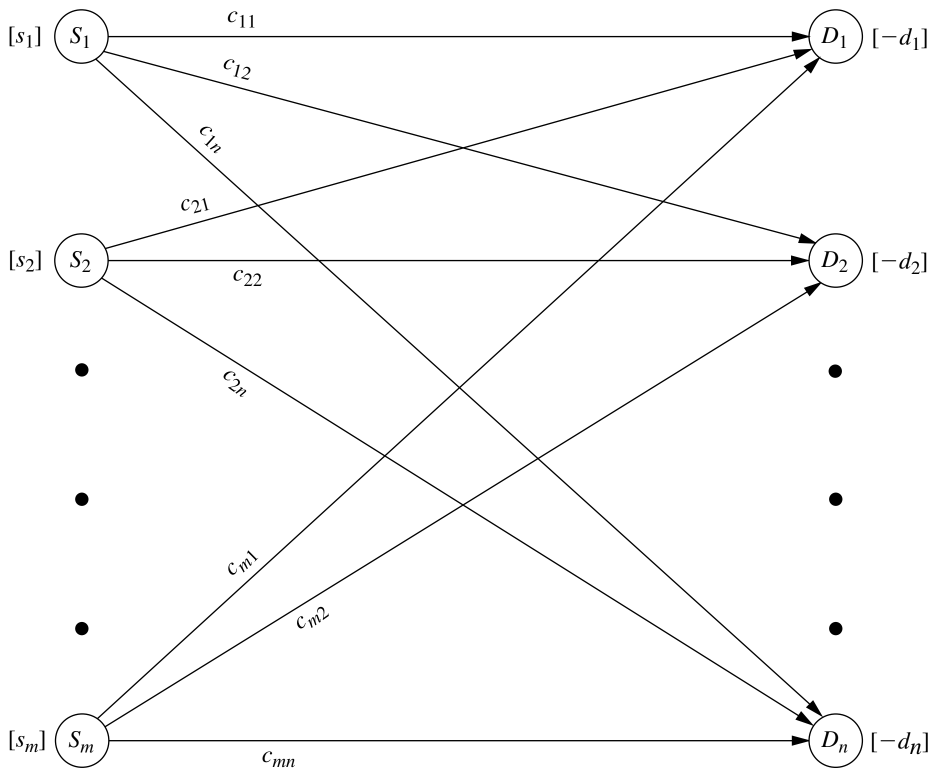
y

$$x_{ij} \geq 0, \qquad \text{para toda } i \text{ y } j.$$

Observe que la tabla que se obtuvo a partir de los coeficientes de las restricciones tiene la estructura especial que se muestra en la tabla 8.6. *Cualquier* problema de programación lineal que se ajuste a esta formulación especial es del tipo de problemas de transporte, sin importar su contexto físico. En realidad, han existido numerosas aplicaciones no relacionadas con el transporte que se ajustan a esta estructura especial, como se verá en el siguiente ejemplo. (El problema de asignación descrito en la sección 8.3 es un ejemplo adicional.) Ésta es una de las razones por las que el problema de transporte suele considerarse como uno de los tipos especiales de problemas de programación lineal más importantes.

En muchas aplicaciones, las cantidades de abastecimiento o recursos y de demanda (las s_i y las d_j) tienen valores enteros, por lo cual, al trabajar con el modelo, se requerirá que las cantidades distribuidas (las x_{ij}) tomen también valores enteros. Por fortuna, gracias a la estructura especial que se muestra en la tabla 8.6, todos los problemas de este tipo tienen la siguiente propiedad.

Propiedad de soluciones enteras: En los casos de problemas de transporte en los que s_i y d_j toman un valor entero, todas las variables básicas (asignaciones), de *toda* solución básica factible (inclusive la óptima), asumen también valores *enteros*.



■ **FIGURA 8.3**
Representación de red del problema de transporte.

El procedimiento de solución que se describe en la sección 8.2 maneja sólo soluciones BF; en consecuencia, en este caso obtendrá, de manera automática, una solución óptima *entera*. (Se podrá ver por qué esta solución proporciona una demostración de la propiedad de soluciones enteras una vez que se conozca el procedimiento; el problema 8.2-20 es una guía para el razonamiento.) Por tanto, no es necesario agregar al modelo la restricción de que las x_{ij} deben tomar sólo valores enteros.

Igual que en el caso de los otros problemas de programación lineal, se dispone de las opciones usuales de software —Excel, LINGO/LINDO, MPL/CPLEX— para establecer y resolver problemas de transporte (y asignación), como se puede ver en los archivos de este capítulo en el OR Courseware. Sin embargo, como el enfoque de Excel es un poco diferente a los que se ha visto hasta ahora, éste se describirá a continuación.

Uso de Excel para formular y resolver problemas de transporte

Como se describió en la sección 3.6, el proceso de usar una hoja de cálculo para formular un modelo de programación lineal para un problema comienza por definir las respuestas a tres preguntas. ¿Cuál

■ **TABLA 8.6** Tabla de coeficientes de las restricciones del problema de transporte

	Coeficiente de:													
	x_{11}	x_{12}	⋯	x_{1n}	x_{21}	x_{22}	⋯	x_{2n}	⋯	x_{m1}	x_{m2}	⋯	x_{mn}	
A =	1	1	⋯	1										Restricciones de suministro
					1	1	⋯	1						
									⋱					
										1	1	⋯	1	
	1				1					1				Restricciones de demanda
		1				1			⋯		1			
			⋱				⋱					⋱		
				1				1					1	

es la *decisión* que se tomará? ¿Cuáles son las *restricciones* sobre estas decisiones? ¿Cuál es la *medida global de desempeño* de estas decisiones? Como el problema de transporte es un tipo especial de problema de programación lineal, abordar estas preguntas también es un punto inicial adecuado para formular este tipo de problemas en una hoja de cálculo. Por tanto, el diseño de la hoja de cálculo gira en torno de configurar esta información y los datos asociados en una forma lógica.

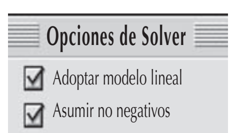
A manera de ilustración, considere de nuevo el problema de la P & T Co. Las decisiones que deberán tomarse están relacionadas con el número de camiones cargados de chícharos que serán enviados de cada enlatadora a cada centro de distribución. Las restricciones sobre estas decisiones son que la cantidad total enviada desde cada enlatadora debe ser igual a su producción (el suministro) y la cantidad total recibida en cada centro de distribución debe ser igual a su asignación (la demanda). La medida global de desempeño es el costo total de envío, por lo que el objetivo es minimizar esta cantidad.

Esta información conduce al modelo en hoja de cálculo que se muestra en la figura 8.4. Todos los datos proporcionados en la tabla 8.2 se despliegan en las siguientes celdas de datos: CostoUnitario (D5:G7), Suministro (J12:J14) y Demanda (D17:G17). Las decisiones sobre las cantidades enviadas son proporcionadas por las celdas cambiantes CantidadDeEnvíos (D12:G14). Las celdas

FIGURA 8.4
Formulación en hoja de cálculo del problema de la P & T Co. como uno de transporte; se incluye la celda objetivo CostoTotal (J17) y las otras celdas de salida TotalEnviado (H12:H14) y TotalRecibido (D15:G15), así como las especificaciones necesarias para preparar el modelo. Las celdas cambiantes CantidadDeEnvíos (D12:G14) muestran el plan de envíos óptimo que se obtuvo mediante el Solver.

	A	B	C	D	E	F	G	H	I	J
1	**Problema de distribución de P&T Co.**									
2										
3		**Costo uni.**		Origen (enlatadora)						
4				Sacramento	Salt Lake City	Rapid City	Albuquerque			
5		Destino	Bellingham	$464	$513	$654	$867			
6		(almacén)	Eugene	$352	$416	$690	$791			
7			Albert Lea	$995	$682	$388	$685			
8										
9										
10		**Cantidad por enviar**		Destino (almacén)						
11		**(Cargas de camión)**		Sacramento	Salt Lake City	Rapid City	Albuquerque	Total recibido		Suministro
12		Origen	Bellingham	0	20	0	55	75	=	75
13		(enlatadora)	Eugene	80	45	0	0	125	=	125
14			Albert Lea	0	0	70	30	100	=	100
15			Total enviado	80	65	70	85			
16				=	=	=	=			Costo total
17			Demanda	80	65	70	85			$ 152 535





	H
11	Total enviado
12	=SUMA(D12:G12)
13	=SUMA(D13:G13)
14	=SUMA(D14:G14)

	C	D	E	F	G
15	Total recibido	=SUMA(D12:D14)	=SUMA(E12:E14)	=SUMA(F12:F14)	=SUMA(G12:G14)

Nombre de rango	Celdas
Demanda	D17:G17
CantidadDeEnvíos	D12:G14
Suministro	J12:J14
CostoTotal	J17
TotalRecibido	D15:G15
TotalEnviado	H12:H14
CostoUnitario	D5:G7

	J
16	Costo total
17	=SUMAPRODUCTO(CostoUnitario,CantidadDeEnvíos)

de salida son TotalEnviado (H12:H14) y TotalRecibido (D15:G15), donde las funciones de SUMA que se introducen en estas celdas se muestran cerca del final de la figura 8.4. Las restricciones, TotalEnviado (H12:H14) = Suministro (J12:J14) y TotalRecibido (D15:G15) = Demanda (D17:G17), se han especificado en la hoja de cálculo y se han introducido en el cuadro de diálogo del Solver. La celda objetivo es CostoTotal (J17), donde su función SUMAPRODUCTO se muestra en la esquina inferior derecha de la figura 8.4. El cuadro de diálogo de Solver especifica que el objetivo es minimizar esta celda objetivo. Una de las opciones seleccionadas de Solver (asumir no negativos) especifica que todas las cantidades enviadas deben ser no negativas. La otra (adoptar modelo lineal) indica que este problema de transporte es de programación lineal.

Para comenzar el proceso de solución del problema se puede introducir cualquier valor (como 0) en cada una de las celdas cambiantes. Después de oprimir el botón “resolver”, el Solver usa el método símplex para obtener la solución del problema. La solución óptima que resulta se muestra en CantidadDeEnvíos(D12:G14) en la figura 8.4, junto con el valor resultante, $152 535, en la celda objetivo CostoTotal (J17).

Observe que Solver usa el método símplex general para resolver el problema de transporte en lugar de la versión simplificada de diseño especial para resolver estos problemas con mucha eficiencia, que se presenta en la siguiente sección. Por tanto, un paquete de software que incluya esta versión simplificada debe resolver un problema de transporte grande con mucha mayor rapidez que Solver de Excel.

Hemos mencionado que algunos problemas no se ajustan por completo al modelo de un problema de transporte porque incumplen el supuesto de requerimientos, pero es posible reformular tales problemas para que se adecuen con la introducción de un destino o un origen ficticios. Si se utiliza Excel Solver, no es necesario reformular puesto que el método símplex puede resolver el modelo original cuando las restricciones están en la forma $\leq$ o cuando las restricciones de demanda tienen la forma $\geq$. (Los archivos de Excel para los siguientes dos ejemplos que contiene el OR Courseware presentan formulaciones en hoja de cálculo que conservan restricciones de suministro o de demanda en su forma original de desigualdad.) Sin embargo, mientras más grande sea el problema, más vale la pena hacer la reformulación y usar el método símplex de transporte (o su equivalente) en lugar de otro paquete de computadora.

Los siguientes dos ejemplos ilustran cómo hacer este tipo de reformulación.

Ejemplo con un destino ficticio

La NORTHERN AIRPLANE COMPANY construye aviones comerciales para varias líneas aéreas de todo el mundo. La última etapa del proceso de producción consiste en fabricar las turbinas de jet e instalarlas —una operación muy rápida— en la estructura del avión terminado. La compañía tiene varios contratos de trabajo que la obligan a entregar un número considerable de aviones en un futuro cercano y en este momento debe programar la producción de motores de turbina para los próximos cuatro meses.

En la segunda columna de la tabla 8.7 se indica la cantidad de motores que debe estar lista para su instalación a fin de cumplir con las fechas de entrega contratadas. De ella se desprende que el número acumulado de motores que deben producirse al final de los meses 1, 2, 3 y 4 debe ser por lo menos de 10, 25, 50 y 70 unidades, respectivamente.

Las instalaciones disponibles para producir las turbinas varían de acuerdo con otros programas de producción, mantenimiento y renovación durante el periodo. Las diferencias mensuales debidas al número máximo que se puede producir y el costo unitario de producción (en millones de dólares) se presentan en la tercera y cuarta columnas de la tabla 8.7.

Dadas las variaciones de los costos de producción, podría valer la pena fabricar algunas turbinas uno o más meses antes de su fecha de instalación; en la actualidad se estudia esta posibilidad. El inconveniente es que esas turbinas deberán almacenarse hasta que sean instaladas, pues la estructura de los aviones no estará lista antes. El costo de almacenamiento de cada turbina es de 15 mil dólares por mes —suma que incluye el interés sobre el capital invertido—,[1] como se muestra en la última columna de la tabla 8.7.

[1] Para propósitos de modelado, suponga que se incurre en este costo de almacenaje al *final del mes* sólo en relación con aquellas unidades que se guardan durante todo el mes hasta el siguiente. Se supondrá que las turbinas producidas durante un mes determinado para instalarlas en ese mismo mes no causan costos de almacenamiento.

■ **TABLA 8.7** Datos de programación de la producción de la Northern Airplane Co.

Mes	Instalaciones programadas	Producción máxima	Costo unitario de producción*	Costo unitario de almacenaje*
1	10	25	1.08	0.015
2	15	35	1.11	0.015
3	25	30	1.10	0.015
4	20	10	1.13	

*El costo está expresado en millones de dólares.

El gerente de producción quiere desarrollar la programación del número de turbinas que se deben fabricar en cada uno de los cuatro meses, de manera que se minimicen los costos totales de producción y almacenamiento.

Formulación. Una manera de formular un modelo matemático para este problema es definir x_j como el número de turbinas que se producen en el mes j, para $j = 1, 2, 3, 4$. Con sólo estas cuatro variables de decisión, se puede formular el problema como un modelo de programación lineal que *no* se ajusta al tipo del problema de transporte (vea el problema 8.2-18).

Por otro lado, si se adopta un punto de vista diferente, sin duda se puede plantear como un problema de transporte cuya solución requiere *mucho* menos esfuerzo. Este punto de vista diferente describirá el problema en términos de orígenes y destinos, y después identificará las x_{ij}, c_{ij}, s_i y d_j correspondientes. (Intente hacerlo antes de leer lo que sigue.)

Como las unidades que se van a distribuir son las turbinas, cada una de las cuales debe programarse para un mes en particular e instalarse en un mes específico —tal vez un mes distinto—,

$$\begin{aligned}
\text{Origen } i &= \text{producción de turbinas en el mes } i \; (i = 1, 2, 3, 4)\\
\text{Destino } j &= \text{instalación de turbinas en el mes } j \; (j = 1, 2, 3, 4)\\
x_{ij} &= \text{número de turbinas producidas en el mes } i \text{ para instalarlas en el mes } j\\
c_{ij} &= \text{costo asociado con cada unidad de } x_{ij}\\
&= \begin{cases} \text{costo por unidad de producción y cualquier almacenaje} & \text{si } i \le j \\ ? & \text{si } i > j \end{cases}\\
s_i &= ?\\
d_j &= \text{número de instalaciones programadas en el mes } j.
\end{aligned}$$

La tabla (incompleta) correspondiente, con los costos y requerimientos se proporciona en la tabla 8.8. Lo que resta por hacer es identificar las cifras de costos y suministros que faltan.

Como es imposible producir turbinas en un mes determinado para instalarlas el mes anterior, x_{ij} debe ser cero para $i > j$. Por tanto, no hay un costo real que se pueda asociar a esas x_{ij}. De cualquier

■ **TABLA 8.8** Tabla incompleta de parámetros de la Northern Airplane Co.

		Costo por unidad distribuida				
		Destino				
		1	2	3	4	Recurso
Origen	1	1.080	1.095	1.110	1.125	?
	2	?	1.110	1.125	1.140	?
	3	?	?	1.100	1.115	?
	4	?	?	?	1.130	?
Demanda		10	15	25	20	

modo, con el objeto de tener un problema de transporte bien definido al que se le pueda aplicar un paquete de computadora estándar —procedimiento de solución de la sección 8.2—, es necesario asignar algún valor a los costos no identificados. Por fortuna, se puede usar el método de la gran M que se introdujo en la sección 4.6 para asignar este valor. Así, se asigna un número *muy* grande —que por conveniencia se denota por M— a estos costos no identificados en la tabla de costos para forzar a que los valores correspondientes de x_{ij} sean iguales a cero en la solución final.

Los números que deben insertarse en la columna de los recursos de la tabla 8.8 no son obvios, porque esos "suministros", es decir, las cantidades producidas en los meses respectivos, no son cantidades fijas. En realidad, el objetivo es encontrar los mejores valores para estas cantidades. Como ya se ha dicho, es necesario asignar un número fijo a todos los elementos de la tabla, incluso a los de la columna de suministros, para tener un problema de transporte. Aunque las restricciones sobre el abastecimiento no estén presentes en la forma usual, existen en forma de cotas superiores sobre la cantidad que se puede producir, es decir,

$$x_{11} + x_{12} + x_{13} + x_{14} \leq 25,$$
$$x_{21} + x_{22} + x_{23} + x_{24} \leq 35,$$
$$x_{31} + x_{32} + x_{33} + x_{34} \leq 30,$$
$$x_{41} + x_{42} + x_{43} + x_{44} \leq 10.$$

El único cambio respecto del modelo estándar del problema de transporte es que estas restricciones están en forma de desigualdades en lugar de igualdades.

Para convertir las desigualdades en ecuaciones con el fin de que se ajusten al problema de transporte, se usa la técnica ya conocida de introducir *variables de holgura*, descrita en la sección 4.2. En el contexto de este ejemplo, las variables de holgura se interpretan como asignaciones a un **destino ficticio** que representa la *capacidad de producción no utilizada* en los meses correspondientes. Este ajuste permite que el abastecimiento, en la formulación del problema de transporte, sea la capacidad de producción total en ese mes. Más aún, como la demanda del destino ficticio es la capacidad total no utilizada, esta demanda es

$$(25 + 35 + 30 + 10) - (10 + 15 + 25 + 20) = 30.$$

Con esta demanda incluida, la suma de los suministros ahora debe ser igual a la suma de las demandas, que es la condición dada por la *propiedad de soluciones factibles* para que éstas existan.

Los datos de costo asociados con el destino ficticio deben ser iguales a cero, ya que la asignación ficticia no significa costo alguno. (Los costos con valor M serían *inapropiados* para esta columna, pues no se trata de forzar que los valores correspondientes de x_{ij} sean iguales a cero. En realidad, esos valores deben sumar 30.)

En la tabla 8.9 se proporciona la tabla final de costos y requerimientos en la que el destino ficticio se indica como 5(D). Cuando se emplea esta formulación y el procedimiento de solución que se describe en la sección 8.2, es bastante fácil encontrar el programa de producción óptimo (vea el problema 8.2-10 y su respuesta al final del libro).

TABLA 8.9 Tabla completa de parámetros de la Northern Airplane Co.

		Costo por unidad distribuida					
		Destino					
		1	2	3	4	5(*D*)	Recursos
Origen	1	1.080	1.095	1.110	1.125	0	25
	2	*M*	1.110	1.125	1.140	0	35
	3	*M*	*M*	1.100	1.115	0	30
	4	*M*	*M*	*M*	1.130	0	10
Demanda		10	15	25	20	30	

Ejemplo con un origen ficticio

METRO WATER DISTRICT es una dependencia que administra la distribución de agua en cierta región geográfica grande. La región es bastante árida, por lo que el distrito debe comprar y traer agua del exterior. Las fuentes de esta agua importada son los ríos Colombo, Sacron y Calorie. El distrito revende el agua a los usuarios de la región. Sus clientes principales son los departamentos de aguas de las ciudades de Berdoo, Los Devils, San Go y Hollyglass.

Es posible hacer llegar agua a cualquiera de estas ciudades desde cualquiera de los tres ríos, con la excepción de que no hay forma de abastecer a Hollyglass con agua del río Calorie. Sin embargo, dada la distribución geográfica de los acueductos y las ciudades en la región, el costo del abastecimiento del distrito depende tanto de la fuente como de la ciudad a la que abastece. En la tabla 8.10 se dan los costos variables por acre-pie de agua —en decenas de dólares— para cada combinación de río y ciudad. A pesar de estas variaciones, el precio que el distrito cobra por acre-pie es independiente de la fuente de agua y es el mismo para todas las ciudades.

La administración del distrito tiene que resolver el problema de cómo asignar el agua disponible durante el próximo verano. En la columna del lado derecho de la tabla 8.10 se proporcionan las cantidades disponibles en los tres ríos, en unidades de un millón de acres-pie. El distrito se compromete a proporcionar una cantidad mínima para cumplir con las necesidades esenciales de cada ciudad —con la excepción de San Go, que tiene una fuente independiente de agua—; estas *necesidades mínimas* se muestran en el renglón correspondiente de la tabla. El renglón de *solicitado* indica que Los Devils no quiere más agua que la que cubre sus necesidades mínimas, pero Berdoo compraría hasta 20 más, San Go hasta 30 más y Hollyglass compraría toda la que pudiera obtener.

La administración desea asignar *toda* el agua disponible de los tres ríos de manera que por lo menos se cumpla con las necesidades mínimas de cada ciudad y al mismo tiempo se minimice el costo total para el distrito.

Formulación. La tabla 8.10 ya casi está en la forma apropiada de una tabla de parámetros, en donde los ríos son los orígenes y las ciudades, los destinos. La única dificultad básica es que no ha quedado claro cuáles deben ser las demandas en cada destino. En realidad, la cantidad que debe recibirse en cada una —excepto en Los Devils— es una variable de decisión con cotas superior e inferior. La cota superior es la cantidad solicitada a menos que exceda la cantidad total disponible después de cumplir con las necesidades mínimas de las otras ciudades, en cuyo caso, esta *cantidad disponible* se convierte en la cota superior. La sedienta ciudad de Hollyglass tiene una cota superior de

$$(50 + 60 + 50) - (30 + 70 + 0) = 60.$$

Sin embargo, las cantidades de las demandas que se incluyen en la tabla de parámetros de un problema de transporte deben ser *constantes*, no variables de decisión acotadas. Para evitar esta dificultad, suponga de momento que no es necesario satisfacer las necesidades mínimas con agua de estos ríos, de manera que las cotas superiores son las únicas restricciones sobre las cantidades que deben asignarse. En estas circunstancias, la pregunta es si las cantidades solicitadas pueden tomarse como la demanda en la formulación de un problema de transporte. ¡La respuesta es sí!, después de hacer un ajuste. (¿Ya se dio cuenta el lector de cuál es el ajuste necesario?)

La situación es análoga al problema de programación de la producción de la Northern Airplane Co., donde existía un *exceso de capacidad de abastecimiento* pero ahora se tiene un *exceso de capacidad de demanda*. En consecuencia, en lugar de introducir un *destino ficticio* para "recibir" los recursos que sobran, el ajuste será introducir un **origen ficticio** para "enviar" la *capacidad de*

TABLA 8.10 Datos de recursos de agua del Metro Water District

	Costo (en decenas de dólares) por acre-pie				
	Berdoo	Los Devils	San Go	Hollyglass	Recursos
Río Colombo	16	13	22	17	50
Río Sacron	14	13	19	15	60
Río Calorie	19	20	23	—	50
Mínimo necesario	30	70	0	10	(en millones de
Solicitado	50	70	30	∞	acres-pie)

■ **TABLA 8.11** Tabla de parámetros sin las necesidades mínimas del Metro Water District

		Costo (decenas de millones de dólares) por unidad distribuida				
		Destino				
		Berdoo	Los Devils	San Go	Hollyglass	Recursos
Origen	Río Colombo	16	13	22	17	50
	Río Sacron	14	13	19	15	60
	Río Calorie	19	20	23	M	50
	Ficticio	0	0	0	0	50
Demanda		50	70	30	60	

demanda no utilizada. La cantidad imaginaria de recursos para este origen ficticio corresponde al excedente de la suma de las demandas sobre la suma de los recursos reales:

$$(50 + 70 + 30 + 60) - (50 + 60 + 50) = 50.$$

Este planteamiento conduce a la tabla de parámetros 8.11, en la que las unidades empleadas son millones de acres-pie y decenas de millones de dólares. Los costos en el renglón *ficticio* son cero porque las asignaciones ficticias no cuestan. Por otro lado, se asigna un costo alto, M, al elemento del río Calorie a Hollyglass. La razón es que el agua del río Calorie no se puede usar para alimentar a la ciudad de Hollyglass y al asignar el costo M se evita esta asignación.

Ahora se analizará cómo se pueden tomar en cuenta las necesidades mínimas de cada ciudad en este tipo de formulaciones. Como San Go no estableció una necesidad mínima, no hay nada que hacer. De forma parecida, la formulación para Hollyglass no requiere ajuste porque su demanda (60) excede en 10 la cantidad disponible en el origen ficticio (50), con lo que la cantidad que le debe llegar desde los orígenes *reales* será *por lo menos de 10* en cualquier solución factible y, por tanto, su necesidad mínima de 10 queda garantizada. (Si no hubiera ocurrido esta coincidencia, tendría que hacerse el mismo ajuste que para Berdoo.)

La necesidad mínima de Los Devils es igual a la cantidad solicitada, así que debe satisfacerse su demanda *completa* de 70 con agua de los orígenes reales y no del ficticio. Esta restricción necesita la ayuda del método de la gran M. Si se asigna un costo muy alto de M a la asignación del origen ficticio a Los Devils, se asegura que esta asignación sea igual a cero en la solución óptima.

Por último, considere Berdoo. A diferencia de Hollyglass, los recursos del origen ficticio son suficientes para "abastecer" —en forma imaginaria— por lo menos parte de la necesidad mínima de Berdoo además de la cantidad adicional que solicita. Entonces, si la necesidad mínima es 30, se deben hacer ajustes para evitar que el origen ficticio contribuya en más de 20 al abastecimiento total (50) de Berdoo. Este objetivo se logra si se divide Berdoo en dos destinos, uno con 30 unidades de demanda y un costo M para cualquier asignación que provenga del origen ficticio, y la otra con una demanda de 20 y con costo unitario de cero para la asignación del destino ficticio. Con esta formulación se obtiene la tabla final de parámetros 8.12.

■ **TABLA 8.12** Tabla de costos y requerimientos del Metro Water District

			Costo (decenas de millones de dólares) por unidad distribuida					
			Destino					
			Berdoo (mín.) 1	Berdoo (adicional) 2	Los Devils 3	San Go 4	Hollyglass 5	Recursos
Origen	Río Colombo	1	16	16	13	22	17	50
	Río Sacron	2	14	14	13	19	15	60
	Río Calorie	3	19	19	20	23	M	50
	Ficticio	4(*D*)	M	0	M	0	0	50
Demanda			30	20	70	30	60	

Este problema se resolverá en la sección 8.2.

Generalizaciones del problema de transporte

Aun después de los tipos de reformulaciones que se presentaron en los dos ejemplos anteriores, algunos problemas que involucran la distribución de unidades de orígenes a destinos no satisfacen el modelo de un problema de transporte. Una razón puede ser que la distribución no va de manera directa de los orígenes a los destinos sino que pasa a través de puntos de transferencia a lo largo del camino. El ejemplo de la Distribution Unlimited Co., de la sección 3.4 (vea la figura 3.13) ilustra tal problema. En este caso, los orígenes son dos fábricas y los destinos son dos almacenes. Sin embargo, un envío desde una fábrica en particular a un almacén específico puede primero ser transferido a un centro de distribución, o incluso a la otra fábrica o al otro almacén antes de llegar a su destino. Los costos unitarios de envío difieren para estos canales de envío diferentes. Aún más, existen límites superiores sobre cuánto se puede embarcar a través de los distintos canales de envío. Aunque éste no es un problema de transporte, todavía es un tipo especial de problema de programación lineal, llamado *problema de flujo de costo mínimo*, que se estudiará en la sección 9.6. El *método símplex de red* que se describe en la sección 9.7 proporciona un modo eficiente de resolver problemas de flujo de costo mínimo. Un problema de este tipo que no impone ningún límite superior sobre cuánto se puede embarcar a través de los canales de envío, se conoce como *problema de transferencia*. La sección 23.1 en el sitio de internet de este libro está dedicada al estudio de esta clase de problemas.

En otros casos, la distribución puede ir en forma directa de las fuentes a los destinos, pero se pueden incumplir otros supuestos del problema de transporte. El *supuesto de costo* se violará si el costo de distribuir unidades de algún origen particular a algún destino específico es una función no lineal del número de unidades distribuidas. El *supuesto de requerimientos* se violará si los recursos de los orígenes o las demandas de los destinos no son fijos. Por ejemplo, la demanda final de un destino puede no conocerse hasta después de que las unidades hayan llegado y entonces se incurre en un costo no lineal cuando la cantidad recibida difiere de la demanda final. Si el suministro de un origen no está fijo, el costo de producir la cantidad suministrada puede ser una función no lineal de esta cantidad. Por ejemplo, un costo fijo puede ser parte del costo asociado con una decisión de abrir un origen nuevo. Se ha realizado una considerable cantidad de investigaciones para generalizar el problema de transporte y su procedimiento de solución en estas situaciones.[2]

8.2 MÉTODO SÍMPLEX MEJORADO PARA SOLUCIONAR EL PROBLEMA DE TRANSPORTE

El problema de transporte es sólo un tipo especial de problemas de programación lineal y puede resolverse con el método símplex como se describió en el capítulo 4. Sin embargo, en esta sección se verá que, si se aprovecha la estructura especial que se muestra en la tabla 8.6, se puede lograr un importante ahorro de cálculos. Se hará referencia a este procedimiento mejorado como el **método símplex de transporte**.

Al avanzar en la lectura, observe en particular la forma en que se aprovecha la estructura especial para desarrollar un menor esfuerzo computacional. Este ahorro ilustra una técnica importante de IO: la simplificación de un algoritmo para aprovechar la estructura especial del problema que se tiene que resolver.

Preparación para el método símplex de transporte

Para resaltar la simplificación lograda por el método símplex de transporte, se revisará primero la forma en que el método símplex general (no simplificado) establecería el problema de transporte en forma tabular. Después de construir la tabla de los coeficientes de restricción (vea la tabla 8.6), de convertir la función objetivo a la forma de maximización y de usar el método de la gran M para introducir las variables artificiales $z_1, z_2, \ldots, z_{m+n}$ en las $m + n$ ecuaciones de restricción respectivas (sección 4.6), se observa que las columnas de la tabla símplex tendrían la forma que se muestra en la tabla 8.13, donde todos los elementos que *no se muestran* en estas columnas son iguales a *cero*.

[2] Por ejemplo, vea K. Holmberg y H. Tuy, "A Production-Transportation Problem with Stochastic Demand and Concave Production Costs", en *Mathematical Programming Series A*, **85:** 157-179, 1999.

TABLA 8.13 Tabla símplex original antes de aplicar el método símplex al problema de transporte

Variable básica	Ec.	Coeficiente de:								Lado derecho
		Z	$\cdots$	x_{ij}	$\cdots$	z_i	$\cdots$	z_{m+j}	$\cdots$	
Z	(0)	-1		c_{ij}		M		M		0
	(1)									
	$\vdots$									
z_i	(i)	0		1		1				s_i
	$\vdots$									
z_{m+j}	$(m+j)$	0		1				1		d_j
	$\vdots$									
	$(m+n)$									

[El único ajuste que queda por hacer antes de la primera iteración es eliminar en forma algebraica los coeficientes distintos a cero de las variables básicas iniciales (artificiales) del renglón 0.]

Después de cualquier iteración subsecuente, el renglón 0 tendría la forma que se muestra en la tabla 8.14. Debido al patrón de ceros y unos que siguen los coeficientes en la tabla 8.13, la *idea fundamental* que se presentó en la sección 5.3 implica que u_i y v_j tienen la siguiente interpretación:

$u_i =$ múltiplo del renglón *i original* que se restó (de manera directa o indirecta) del renglón 0 *original* en todas las iteraciones del método símplex que condujeron a la tabla actual.

$v_j =$ múltiplo del renglón $m + j$ *original* que se restó (en forma directa o indirecta) del renglón 0 *original* en todas las iteraciones del método símplex que condujeron a la tabla actual.

Si se aplica la teoría de dualidad expuesta en el capítulo 6, se puede reconocer que las u_i y v_j son las *variables duales*.[3] Si x_{ij} es una variable no básica, $c_{ij} - u_i - v_j$ se interpreta como la tasa a la que cambiaría Z si se aumentara el valor de x_{ij}.

La información necesaria. Con el fin de establecer las bases que permitan la simplificación, recuerde cuál es la información que necesita el método símplex. En el paso inicial, debe obtenerse una solución BF inicial, lo que se hace en forma artificial mediante la introducción de variables artificiales para que constituyan el conjunto de variables básicas iniciales cuyo valor es igual a las s_i y d_j. Para llevar a cabo la prueba de optimalidad y el paso 1 de una iteración —seleccionar una variable básica entrante— se requiere conocer el renglón 0 actual, que se obtiene al restar del renglón 0 anterior algún múltiplo de otro renglón. El paso 2 —determinación de la variable básica que sale— debe identificar la variable básica que llega primero a cero cuando aumenta el valor de la variable básica entrante; esto se hace al comparar los coeficientes actuales de la variable básica entrante con el lado derecho correspondiente. El paso 3 debe determinar la nueva solución BF, que se encuentra al restar algunos múltiplos de un renglón, de otros renglones de la tabla símplex actual.

Formas muy simplificadas para obtener esta información. Ahora, ¿cómo obtiene el *método símplex de transporte* la misma información de una manera mucho más sencilla? La respuesta a esta pregunta es el tema de las páginas siguientes, pero se dará alguna información preliminar.

Primero, *no se necesitan variables artificiales*, puesto que se dispone de un procedimiento sencillo y conveniente —con algunas variaciones— para construir una solución inicial básica factible.

Segundo, el renglón 0 actual se puede obtener *sin usar ningún otro renglón* con sólo calcular los valores de u_i y v_j de manera directa. Como cada variable básica debe tener coeficiente igual a cero en el renglón 0, estos valores de u_i y v_j se pueden obtener si se resuelve el sistema de ecuaciones

$$c_{ij} - u_i - v_j = 0 \qquad \text{para cada } i \text{ y } j \text{ tal que } x_{ij} \text{ es variable básica.}$$

[3] Será más sencillo reconocer estas variables como duales si se reetiquetan todas como y_i y después se cambian todos los signos del renglón 0 de la tabla 8.14, lo cual convertirá la función objetivo de nuevo a su forma original de minimización.

TABLA 8.14 Renglón 0 de la tabla símplex si se aplica el método símplex al problema de transporte

Variable básica	Ec.	Coeficiente de:								Lado derecho
		Z	⋯	x_{ij}	⋯	z_i	⋯	z_{m+j}	⋯	
Z	(0)	-1		$c_{ij} - u_i - v_j$		$M - u_i$		$M - v_j$		$-\sum_{i=1}^{m} s_i u_i - \sum_{j=1}^{n} d_j v_j$

(Este procedimiento directo se ilustrará más adelante cuando se analice la prueba de optimalidad del método símplex de transporte.) La estructura especial de la tabla 8.13 permite esta forma conveniente de obtener el renglón 0 al proporcionar como coeficientes de x_{ij}, $c_{ij} - u_i - v_j$ en la tabla 8.14.

Tercero, la variable básica que sale se puede identificar de manera sencilla sin usar (en forma explícita) los coeficientes de la variable básica entrante. Esto también se debe a la estructura especial del problema, que facilita ver cómo debe cambiar la solución cuando crece el valor de la variable entrante. Como resultado, la nueva solución BF también se puede identificar de inmediato *sin manipulaciones algebraicas* sobre los renglones de la tabla símplex. (Los detalles se verán más adelante cuando se describa la forma como el método símplex de transporte realiza una iteración.)

La gran conclusión es que *se puede eliminar casi toda la tabla símplex* (junto con todo el esfuerzo de actualizarla). Además de los datos de entrada (los valores de c_{ij}, s_i y d_j), la única información que necesita el método símplex de transporte es la solución BF actual,[4] los valores actuales de u_i y v_j, y los valores resultantes de $c_{ij} - u_i - v_j$ de las variables no básicas x_{ij}. Cuando se resuelve un problema a mano es conveniente registrar esta información en una **tabla símplex de transporte**, como la que se muestra en la tabla 8.15. (Observe con cuidado que los valores de x_{ij} y $c_{ij} - u_i - v_j$ se distinguen en esta tabla, pues encierran en un círculo los primeros pero no los últimos.)

La gran mejora en eficiencia resultante. Es posible apreciar de manera global la gran diferencia en eficiencia y conveniencia entre los métodos símplex de transporte y símplex si ambos se aplican al mismo problema pequeño (vea el problema 8.2-17). Sin embargo, la diferencia es más marcada en problemas grandes que tienen que resolverse en una computadora. El tamaño de las tablas símplex y de transporte sugiere, de alguna manera, esta gran ganancia. En el caso de un problema de transporte que tiene m orígenes y n destinos, la tabla símplex tiene $m + n + 1$ renglones y $(m + 1)(n + 1)$ columnas —sin incluir las que se colocan a la izquierda de las columnas de x_{ij}—, y la tabla símplex de transporte tendrá m renglones y n columnas (se excluyen los dos renglones y columnas informativas adicionales). Si se dan valores a m y n —por ejemplo, $m = 10$ y $n = 100$, que sería un problema bastante normal de tamaño mediano—, se puede observar cómo crece la razón del número de celdas en la tabla símplex entre el mismo número en la tabla de transporte, conforme m y n crecen.

Paso inicial

Recuerde que el objetivo de este paso es obtener una solución inicial BF. Como todas las restricciones funcionales de los problemas de transporte son *igualdades*, el método símplex obtendría esta solución mediante la introducción de variables artificiales y su utilización como variables básicas iniciales, según se describió en la sección 4.6. De hecho, la solución básica que resulta sólo es factible para la versión revisada del problema, por lo que se necesita un buen número de iteraciones para hacer que el valor de estas variables sea igual a cero y se obtengan las soluciones BF reales. El método símplex de transporte evita todo esto, puesto que usa un procedimiento más sencillo para construir de manera directa una solución BF real en la tabla símplex de transporte.

[4] Como las variables no básicas tienen valor cero de manera automática, la solución BF actual se identifica por completo al registrar sólo los valores de las variables básicas. En lo sucesivo se usará esta convención.

TABLA 8.15 Formato de la tabla símplex de transporte

		Destino					
		1	**2**	⋯	***n***	**Recursos**	u_i
Origen	1	c_{11}	c_{12}	⋯	c_{1n}	s_1	
	2	c_{21}	c_{22}	⋯	c_{2n}	s_2	
	⋮	⋯	⋯	⋯	⋯	⋮	
	m	c_{m1}	c_{m2}	⋯	c_{mn}	s_m	
Demanda		d_1	d_2	⋯	d_n	$Z =$	
	v_j						

Información adicional que se agrega en cada celda:

Si x_{ij} es una variable básica	*Si x_{ij} es una variable no básica*
c_{ij} / (x_{ij}) encerrado en un círculo	c_{ij} / $c_{ij} - u_i - v_j$

Antes de describir este procedimiento, es necesario establecer que el número de variables básicas de cualquier solución básica de un problema de transporte es una menos de lo que se espera. Por lo general, en los problemas de programación lineal, se tiene una variable básica por cada restricción funcional. En los problemas de transporte con m recursos y n destinos, el número de restricciones funcionales es $m + n$. Sin embargo,

Número de variables básicas $= m + n - 1$.

Este resultado se debe a las restricciones de igualdad manejadas y a que este conjunto de $m + n$ ecuaciones tiene una ecuación *extra* (o *redundante*) que se puede eliminar sin cambiar la región factible, esto es, cualquiera de las restricciones se satisface en forma automática siempre que las otras $m + n - 1$ restricciones queden satisfechas. (Es posible verificar esta conclusión si se demuestra que cualquiera de las restricciones de recursos es exactamente igual a la suma de las restricciones de demanda menos la suma de las otras restricciones de recursos, y que cualquier restricción de demanda también puede expresarse al sumar las ecuaciones de recursos y restar las otras ecuaciones de demanda. Vea el problema 8.2-19). Por tanto, cualquier *solución BF* en una tabla de transporte debe aparecer con exactamente $m + n - 1$ asignaciones *no negativas* encerradas en un círculo, en donde la suma de las asignaciones en cada renglón o columna es igual a su capacidad o a su demanda.[5]

El procedimiento para construir una solución inicial BF selecciona, una por una, las $m + n - 1$ variables básicas. Después de cada selección, se asigna a esa variable un valor que satisfará una más de las restricciones, de tal forma que se elimina el renglón o columna de esa restricción para cualquier nueva asignación. Una vez hechas las $m + n - 1$ elecciones, el resultado es que se ha construido una solución básica completa, de tal manera que se satisfacen todas las restricciones. Se han propuesto varios criterios diferentes para elegir las variables básicas. A continuación se presentan y ejemplifican tres de ellos después de describir el procedimiento general.

[5] Sin embargo, nótese que cualquier solución básica factible con $m + n - 1$ variables distintas de cero *no necesariamente* es una solución básica porque puede ser el promedio ponderado de dos o más soluciones BF degeneradas, esto es, soluciones BF que tienen algunas variables básicas iguales a cero. No debe preocuparse en cuanto a etiquetar tales soluciones como básicas, pues el método símplex de transporte construye sólo soluciones BF legítimas.

Procedimiento general[6] para construir una solución inicial BF. Al inicio, todos los renglones de los orígenes y las columnas de destinos de la tabla símplex de transporte se toman en cuenta para proporcionar una variable básica (asignación).

1. Se selecciona la siguiente variable básica (asignación) entre los renglones y columnas en que todavía se puede hacer una asignación de acuerdo con algún criterio.
2. Se hace una asignación que sea tan grande como para que use el resto de los recursos en ese renglón o la demanda restante en esa columna (lo que resulte menor).
3. Se elimina ese renglón o columna —la que tenía la cantidad más pequeña de los recursos o demandas restantes— para las nuevas asignaciones. (Si el renglón y la columna tienen la misma cantidad de recursos y demanda restantes, de manera arbitraria se elimina el *renglón*. La columna se usará después para proporcionar una variable básica *degenerada,* es decir, una asignación con cero unidades encerradas en un círculo.)
4. Si sólo queda un renglón o una columna dentro de las posibilidades, entonces el procedimiento se termina eligiendo como básicas cada una de las variables *restantes* —es decir, las variables que no han sido elegidas ni eliminadas al quitar su renglón o columna— asociadas con ese renglón o columna que tiene la única asignación posible. De otra manera, se regresa al paso 1.

Criterios alternativos para el paso 1

1. *Regla de la esquina noroeste:* La primera elección es x_{11}, es decir, se comienza en la esquina noroeste de la tabla símplex de transporte. De ahí en adelante, si x_{ij} fue la última variable básica seleccionada, la siguiente elección es $x_{i,j+1}$ —es decir, se mueve una columna a la *derecha*, si quedan recursos en el origen *i*. De otra manera, se elige $x_{i+1,j}$, esto es, se mueve un renglón hacia *abajo*—.

Ejemplo. Para hacer más concreta esta descripción, se ilustrará el procedimiento general en el problema del Metro Water District (vea la tabla 8.12) con el uso de la regla de la esquina noroeste en el paso 1. Como en este caso $m = 4$ y $n = 5$ el procedimiento encontrará una solución inicial BF que tiene $m + n - 1 = 8$ variables básicas.

Como se muestra en la tabla 8.16, la primera asignación es $x_{11} = 30$, lo que completa la demanda de la columna 1 (y la elimina para nuevas asignaciones). La primera iteración deja 20 unidades de recursos restantes en el renglón 1, y después se elige $x_{1,1+1} = x_{12}$ como variable básica. Dado que los recursos restantes no son mayores que la demanda de 20 unidades de la columna 2, se asignan todos estos recursos a $x_{12} = 20$, y se elimina este renglón. (Se opta por eliminar el renglón 1 y no la columna 2 de acuerdo con la instrucción entre paréntesis del paso 3.) Entonces se selecciona $x_{1+1,2} = x_{22}$. Como la demanda de 0 que queda en la columna 2 es menor que los recursos de 60 en el renglón 2, se asigna $x_{22} = 0$ y se elimina la columna 2.

Si se continúa en esta misma forma, llegará el momento en que se obtenga la *solución inicial BF* completa que se muestra en la tabla 8.16, en la que los números dentro de los círculos son los valores de las variables básicas ($x_{11} = 30, \ldots, x_{45} = 50$) y todas las otras variables (x_{13}, etc.) son no básicas iguales a cero. Se agregaron flechas para indicar el orden en que fueron elegidas las variables básicas (asignaciones). El valor de *Z* para esta solución es

$$Z = 16(30) + 16(20) + \cdots + 0(50) = 2\,470 + 10M.$$

2. *Método de aproximación de Vogel:* Para cada renglón y columna que queda bajo consideración, se calcula su **diferencia**, que se define como *la diferencia aritmética entre el costo unitario más pequeño c_{ij} y el que le sigue de los que quedan en ese renglón o columna*. (Si se tiene un empate entre el costo más pequeño de los restantes de un renglón o columna, entonces la *diferencia* es 0.) En el renglón o columna que tiene la *mayor diferencia*, se elige la variable que tiene el *menor costo unitario que queda*. (Los empates en el caso de la mayor de estas diferencias se pueden romper de manera arbitraria.)

[6] En la sección 4.1 se hizo notar que el método símplex es un ejemplo de los algoritmos —procedimientos iterativos de solución— que más se usan en el trabajo de IO. Observe que este procedimiento también es un algoritmo, en el que cada ejecución sucesiva de los (cuatro) pasos constituye una iteración.

TABLA 8.16 Solución inicial BF de acuerdo con la regla de la esquina noroeste

		Destino						
		1	**2**	**3**	**4**	**5**	**Recursos**	u_i
Fuente	1	16 (30)	16 (20)	13	22	17	50	
	2	14	14 (0)	13 (60)	19	15	60	
	3	19	19	20 (10)	23 (30)	M (10)	50	
	4(D)	M	0	M	0	0 (50)	50	
Demanda		30	20	70	30	60	$Z = 2\,470 + 10M$	
	v_j							

Ejemplo. Ahora se aplicará el procedimiento general al problema del Metro Water District con el criterio del método de aproximación de Vogel para seleccionar la siguiente variable básica en el paso 1. Al aplicarlo, es más conveniente trabajar con la tabla de parámetros —en lugar de la tabla símplex de transporte completa—, esto es, se comienza con la que se muestra en la tabla 8.12. En cada iteración, después de calcular y escribir las diferencias en cada renglón y columna que quedan bajo consideración, se encierra en un círculo la mayor de ellas y se enmarca en un cuadro el costo unitario menor en ese renglón o columna. La variable con este costo unitario menor se selecciona como la siguiente variable básica y su valor se indica en la esquina inferior derecha de la tabla actual, junto con el renglón o columna que se elimina (vea los pasos 2 y 3 del procedimiento general). La tabla de la siguiente iteración es la misma, pero se elimina este renglón o columna y se resta la cantidad asignada de la demanda o los recursos correspondientes (cualesquiera que sean los sobrantes).

La aplicación de este procedimiento al problema del Metro Water District resulta en la serie de tablas de parámetros que se muestran en la tabla 8.17, en donde la solución BF inicial consiste en las ocho variables básicas (asignaciones) dadas en la esquina inferior derecha de las tablas de parámetros respectivas.

Este ejemplo ilustra dos características sutiles del procedimiento general que merecen atención especial. Primero, observe que la iteración final elige *tres* variables (x_{31}, x_{32} y x_{33}) para entrar a la base, en lugar de una sola que se elige en las iteraciones anteriores. La razón es que en este punto queda sólo *un* renglón (el renglón 3) sin eliminar. Por tanto, el paso 4 del procedimiento general impone que se deben seleccionar como *básicas* todas las variables restantes asociadas con el renglón 3.

Segundo, observe que la asignación $x_{23} = 20$ en la penúltima iteración agota *tanto* los recursos restantes de ese renglón *como* la demanda que queda en esa columna. Sin embargo, en lugar de eliminar los dos, el paso 3 dispone que se elimine *sólo el renglón* y se deje la columna para que más tarde proporcione una variable básica *degenerada*. En realidad, la columna 3 se usa para este propósito en la iteración final cuando se selecciona $x_{33} = 0$ como una de las variables básicas. Vea en la tabla 8.16 otro ejemplo de este fenómeno, donde después de hacer la asignación $x_{12} = 20$ se elimina sólo el renglón 1, y la columna 2 se conserva para proporcionar la variable básica degenerada $x_{22} = 0$, en la siguiente iteración.

Aunque una asignación de cero puede parecer irrelevante, en realidad juega un papel importante, pues como se verá pronto, el método símplex de transporte debe conocer *todas* las $m + n - 1$ variables básicas, incluso aquellas que tienen un valor de cero, que constituyen la solución BF actual.

■ **TABLA 8.17** **Solución BF inicial de acuerdo con el método de aproximación de Vogel**

		Destino						
		1	2	3	4	5	Recursos	Diferencia por renglón
Fuente	1	16	16	13	22	17	50	3
	2	14	14	13	19	15	60	1
	3	19	19	20	23	*M*	50	0
	4(*D*)	*M*	0	*M*	[0]	0	50	0
Demanda		30	20	70	30	60	Seleccionar $x_{44} = 30$	
Diferencia por columna		2	14	0	(19)	15	Eliminar columna 4	

		Destino					
		1	2	3	5	Recursos	Diferencia por renglón
Fuente	1	16	16	13	17	50	3
	2	14	14	13	15	60	1
	3	19	19	20	*M*	50	0
	4(*D*)	*M*	0	*M*	[0]	20	0
Demanda		30	20	70	60	Seleccionar $x_{45} = 20$	
Diferencia por columna		2	14	0	(15)	Eliminar columna 4(*D*)	

		Destino					
		1	2	3	5	Recursos	Diferencia por renglón
Fuente	1	16	16	[13]	17	50	(3)
	2	14	14	13	15	60	1
	3	19	19	20	*M*	50	0
Demanda		30	20	70	40	Seleccionar $x_{13} = 50$	
Diferencia por columna		2	2	0	2	Eliminar columna 1	

		Destino					
		1	2	3	5	Recursos	Diferencia por renglón
Fuente	2	14	14	13	[15]	60	1
	3	19	19	20	*M*	50	0
Demanda		30	20	20	40	Seleccionar $x_{25} = 40$	
Diferencia por columna		5	5	7	($M - 15$)	Eliminar columna 5	

		Destino				
		1	2	3	Recursos	Diferencia por renglón
Fuente	2	14	14	[13]	20	1
	3	19	19	20	50	0
Demanda		30	20	20	Seleccionar $x_{23} = 20$	
Diferencia por columna		5	5	(7)	Eliminar columna 2	

		Destino				
		1	2	3	Recursos	
Fuente	3	19	19	20	50	
Demanda		30	20	0	Seleccionar $x_{31} = 30$ $x_{32} = 20$ $x_{33} = 0$	$Z = 2\,460$

3. *Método de aproximación de Russell:* Para cada renglón de origen i que queda bajo consideración, debe determinarse $\bar{u}_i$, que es el mayor costo unitario c_{ij} de los que quedan en ese renglón. En el caso de cada columna de destino j que todavía está bajo consideración, se determina $\bar{v}_j$, que es el mayor costo unitario c_{ij} de los que hay en esa columna. Para cada variable x_{ij} que no

haya sido seleccionada en estos renglones o columnas, se calcula $\Delta_{ij} = c_{ij} - \bar{u}_i - \bar{v}_j$. Se elige la variable con el *mayor valor negativo* (en términos absolutos) de Δ_{ij}. (Los empates se pueden romper de manera arbitraria.)

Ejemplo. Se aplicará de nuevo el procedimiento general al problema del Metro Water District (vea la tabla 8.12) con el método de aproximación de Russell. En la tabla 8.18 se muestran los resultados que incluyen la serie de variables básicas (asignaciones).

En la iteración 1, el mayor costo unitario en el renglón 1 es $\bar{u}_1 = 22$, en la columna 1 es $\bar{v}_1 = M$, etc. De esta forma,

$$\Delta_{11} = c_{11} - \bar{u}_1 - \bar{v}_1 = 16 - 22 - M = -6 - M.$$

Al calcular todos los valores de Δ_{ij} para $i = 1, 2, 3, 4$ y $j = 1, 2, 3, 4, 5$ se observa que $\Delta_{45} = 0 - 2M$ tiene el valor negativo mayor, por lo cual se elige $x_{45} = 50$ como la primera variable básica (asignación). Esta asignación agota todos los recursos que se tienen en el renglón 4, por lo que este renglón se elimina.

Observe que esta eliminación cambia $\bar{v}_1$ y $\bar{v}_3$ en la siguiente iteración, por lo que ahora se requiere volver a calcular Δ_{ij} con $j = 1, 3$ al igual que es necesario eliminar $i = 4$. Ahora el valor negativo más grande es

$$\Delta_{15} = 17 - 22 - M = -5 - M,$$

con lo cual $x_{15} = 10$ se convierte en la segunda variable básica (asignación), y se elimina la columna 5.

Las iteraciones siguientes proceden de una manera similar, pero el lector quizá desee probar su comprensión mediante la verificación del resto de las asignaciones dadas en la tabla 8.18. Al igual que con otros procedimientos de esta sección (y de otras) el IOR Tutorial puede resultar útil cuando se trata de hacer cálculos y como ilustración del enfoque. (Vea la rutina interactiva para encontrar una solución BF inicial.)

Comparación de criterios alternativos para el paso 1. A continuación se compararán estos tres criterios para elegir la siguiente variable básica. La virtud principal de la regla de la esquina noroeste es la facilidad y rapidez con que se aplica. Sin embargo, como no le da importancia a los costos unitarios c_{ij}, por lo general la solución que se obtiene distará mucho de la óptima. (Observe en la tabla 8.16 que $x_{35} = 10$ aun cuando $c_{35} = M$.) Si se realiza un esfuerzo un poco mayor para encontrar la solución inicial BF, es posible que se reduzca mucho el número de iteraciones que después necesita el método símplex de transporte para encontrar la solución óptima (vea los problemas 8.2-7 y 8.2-9). El objetivo de los otros dos criterios es precisamente encontrar una solución así.

TABLA 8.18 Solución inicial básica factible de acuerdo con el método de aproximación de Russell

Iteración	$\bar{u}_1$	$\bar{u}_2$	$\bar{u}_3$	$\bar{u}_4$	$\bar{v}_1$	$\bar{v}_2$	$\bar{v}_3$	$\bar{v}_4$	$\bar{v}_5$	Valor negativo más grande de Δ_{ij}	Asignación
1	22	19	M	M	M	19	M	23	M	$\Delta_{45} = -2M$	$x_{45} = 50$
2	22	19	M		19	19	20	23	M	$\Delta_{15} = -5 - M$	$x_{15} = 10$
3	22	19	23		19	19	20	23		$\Delta_{13} = -29$	$x_{13} = 40$
4		19	23		19	19	20	23		$\Delta_{23} = -26$	$x_{23} = 30$
5		19	23		19	19		23		$\Delta_{21} = -24$*	$x_{21} = 30$
6										Irrelevante	$x_{31} = 0$
											$x_{32} = 20$
											$x_{34} = 30$
											$Z = 2\,570$

*El empate con $\Delta_{22} = -24$ se rompe de modo arbitrario.

TABLA 8.19 Tabla símplex de transporte inicial (antes de obtener $c_{ij} - u_i - v_j$) mediante el método de aproximación de Russell

Iteración 0		Destino 1	2	3	4	5	Recursos	u_i
Origen	1	16	16	13 (40)	22	17 (10)	50	
	2	14 (30)	14	13 (30)	19	15	60	
	3	19 (0)	19 (20)	20	23 (30)	*M*	50	
	4(*D*)	*M*	0	*M*	0	0 (50)	50	
Demanda		30	20	70	30	60	$Z = 2\ 570$	
	v_j							

El método de aproximación de Vogel ha sido el más popular durante muchos años,[7] en parte porque es relativamente fácil hacerlo a mano. Este criterio toma en cuenta los costos unitarios en forma eficaz, puesto que la *diferencia* representa el mínimo costo adicional en que se incurre por no hacer una asignación en la celda que tiene el menor costo en esa columna o renglón.

El método de aproximación de Russell proporciona otro criterio excelente[8] que es fácil poner en práctica en una computadora, aunque no a mano. Es cierto que todavía se requiere más experimentación para determinar cuál es más eficiente *en promedio*, pero *con frecuencia* este criterio ha proporcionado una mejor solución que el de Vogel. (En el ejemplo, el método de aproximación de Vogel por casualidad encuentra la solución óptima con $Z = 2\ 460$, mientras que el de Russell falla por muy poco con $Z = 2\ 570$.) En un problema grande, quizá valga la pena aplicar ambos criterios y después usar la mejor solución que se obtenga para iniciar las iteraciones del método símplex de transporte.

Una ventaja peculiar del método de Russell es que sigue el patrón del paso 1 de una iteración del método símplex de transporte —como se verá muy pronto—, lo que de alguna manera simplifica el programa completo de computadora. En particular, los valores de $\overline{u}_i$ y $\overline{v}_j$ se definieron de manera que los valores relativos de $c_{ij} - \overline{u}_i - \overline{v}_j$ son una *estimación* de los valores relativos de $c_{ij} - u_i - v_j$ que se obtendrán cuando el método símplex de transporte alcance la solución óptima.

Ahora se usará la solución inicial básica factible que se obtuvo en la tabla 8.18 por el método de aproximación de Russell para ilustrar el resto del método símplex de transporte. Por ello, en la tabla 8.19 se muestra la *tabla símplex de transporte inicial* (antes de encontrar las u_i y v_j).

El siguiente paso es verificar si esta solución inicial es efectivamente óptima mediante la aplicación de la *prueba de optimalidad*.

Prueba de optimalidad

Si se utiliza la notación de la tabla 8.14, la prueba de optimalidad estándar del método símplex (vea la sección 4.3) para el problema de transporte, se puede reducir como sigue:

> **Prueba de optimalidad:** Una solución BF es óptima si y sólo si $c_{ij} - u_i - v_j \geq 0$ para toda (i, j) tal que x_{ij} es no básica.[9]

[7] N. V. Reinfeld y W. R. Vogel, *Mathematical Programming*, Prentice-Hall, Englewood Cliffs, NJ, 1958.

[8] E. J. Russell, "Extension of Dantzig's Algorithm to Finding an Initial Near-Optimal Basis for the Transportation Problem", en *Operations Research*, **17**: 187-191, 1969.

[9] La única excepción es que dos o más soluciones básicas factibles equivalentes degeneradas —es decir, soluciones idénticas que tienen distintas variables básicas degeneradas iguales a cero— pueden ser óptimas pero sólo una de ellas puede satisfacer la prueba de optimalidad. Esta excepción se ilustra más adelante en el ejemplo (observe las soluciones idénticas en las dos últimas tablas símplex de la tabla 8.23, en donde sólo la última solución satisface el criterio de optimalidad).

Así, lo único que esta prueba requiere es la obtención de los valores de u_i y v_j para la solución básica factible actual y el cálculo de los valores $c_{ij} - u_i - v_j$, como se describe en seguida.

Como el valor de $c_{ij} - u_i - v_j$ debe ser cero si x_{ij} es una variable básica, u_i y v_j satisfacen el conjunto de ecuaciones

$$c_{ij} = u_i + v_j \qquad \text{para cada } (i, j) \text{ tal que } x_{ij} \text{ es básica.}$$

Existen $m + n - 1$ variables básicas y, por tanto, hay $m + n - 1$ ecuaciones de este tipo. Como el número de incógnitas (las u_i y v_j) es $m + n$, se puede asignar un valor arbitrario a cualquiera de estas variables sin incumplir las ecuaciones. La elección de esta variable y su valor no afecta el valor de ningún $c_{ij} - u_i - v_j$, aun cuando x_{ij} sea no básica, por lo que la única diferencia (menor) estriba en la facilidad para resolver estas ecuaciones. Una opción conveniente para lograr esto es seleccionar la u_i que tiene el *mayor número de asignaciones en su renglón* —los empates se rompen de manera arbitraria— y asignarle un valor de cero. Gracias a la estructura sencilla de estas ecuaciones, resulta muy fácil obtener en forma algebraica los valores del resto de las variables.

A manera de demostración, se darán las ecuaciones que corresponden a cada variable básica en la solución BF inicial.

$$\begin{array}{lll}
x_{31}: & 19 = u_3 + v_1. & \text{Se hace } u_3 = 0, \text{ así } v_1 = 19, \\
x_{32}: & 19 = u_3 + v_2. & \qquad\qquad\qquad v_2 = 19, \\
x_{34}: & 23 = u_3 + v_4. & \qquad\qquad\qquad v_4 = 23. \\
x_{21}: & 14 = u_2 + v_1. & \text{Se conoce } v_1 = 19, \text{ así } u_2 = -5. \\
x_{23}: & 13 = u_2 + v_3. & \text{Se conoce } u_2 = -\ 5, \text{ así } v_3 = 18. \\
x_{13}: & 13 = u_1 + v_3. & \text{Se conoce } v_3 = 18, \text{ así } u_1 = -5. \\
x_{15}: & 17 = u_1 + v_5. & \text{Se conoce } u_1 = -5, \text{ así } v_5 = 22. \\
x_{45}: & 0 = u_4 + v_5. & \text{Se conoce } v_5 = 22, \text{ así } u_4 = -22.
\end{array}$$

Si se establece $u_3 = 0$ —puesto que el renglón 3 de la tabla 8.19 tiene el mayor número de asignaciones: 3— y se resuelven una por una las ecuaciones, se obtienen de inmediato los valores de las incógnitas, que se muestran a la derecha. (Observe que obtener los valores de u_i y v_j depende de qué variables x_{ij} son *variables básicas* en la solución BF actual, de manera que será necesario repetir este cálculo cada vez que se obtenga una nueva solución BF.)

Una vez familiarizado con este desarrollo, tal vez encuentre más conveniente resolver las ecuaciones sin escribirlas y trabajar de manera directa sobre la tabla símplex de transporte. Así, en la tabla 8.19 se comienza por escribir el valor de $u_3 = 0$ y después se toma cada una de las asignaciones (x_{31}, x_{32}, x_{34}) de ese renglón, para las que se establece $v_j = c_{3j}$ y en seguida se buscan las asignaciones en estas columnas (excepto en el renglón 3), como (x_{21}). En forma mental se calcula $u_2 = c_{21} - v_1$. Si se toma x_{23}, se establece $v_3 = c_{23} - u_2$, así hasta encontrar y escribir todos los valores de las u_i y v_j (inténtelo). Después se calculan y escriben los valores de $c_{ij} - u_i - v_j$ para cada una de las variables no básicas x_{ij} —es decir, para las celdas sin asignación—; en la tabla 8.20 se muestra la tabla símplex de transporte inicial que se obtiene.

En este momento se puede aplicar la prueba de optimalidad para verificar los valores de $c_{ij} - u_i - v_j$ dados en esta tabla 8.20. Como dos de estos valores, ($c_{25} - u_2 - v_5 = -2$ y $c_{44} - u_4 - v_4 = -1$) son negativos, se concluye que la solución BF actual no es óptima. Entonces, el método símplex de transporte debe proceder a hacer una iteración para encontrar una mejor solución BF.

Una iteración

Igual que en el caso del método símplex estándar, una iteración de esta versión simplificada debe determinar una variable básica entrante (paso 1), una variable básica saliente (paso 2) y después identificar la nueva solución BF que resulta (paso 3).

Paso 1: Encontrar la variable básica entrante. Como $c_{ij} - u_i - v_j$ representa la tasa a la que cambia la función objetivo si se incrementa la variable no básica x_{ij}, la variable básica que entra debe tener un valor de $c_{ij} - u_i - v_j$ *negativo*, para que el costo total Z disminuya. Entonces, los candidatos de la tabla 8.20 son x_{25} y x_{44}. Entre ellos se elige el valor negativo más grande (en términos absolutos) de $c_{ij} - u_i - v_j$ como la variable básica entrante, que en este caso corresponde a x_{25}.

■ **TABLA 8.20** **Tabla símplex de transporte inicial completa**

Iteración 0		**Destino** 1	2	3	4	5	**Recursos**	u_i
Origen	1	[16] +2	[16] +2	[13] (40)	[22] +4	[17] (10)	50	−5
	2	[14] (30)	[14] 0	[13] (30)	[19] +1	[15] −2	60	−5
	3	[19] (0)	[19] (20)	[20] +2	[23] (30)	[*M*] *M* − 22	50	0
	4(*D*)	[*M*] *M* + 3	[0] +3	[*M*] *M* + 4	[0] −1	[0] (50)	50	−22
Demanda		30	20	70	30	60	*Z* = 2 570	
	v_j	19	19	18	23	22		

Paso 2: Encontrar la variable básica saliente. Si se incrementa el valor de la variable básica entrante, se establece una *reacción en cadena* de cambios compensatorios en otras variables básicas (asignaciones) para que se satisfagan las restricciones de recursos y demanda. La primera variable básica que disminuya su valor hasta cero será la variable básica que sale.

Si x_{25} es la variable básica entrante, la reacción en cadena de la tabla 8.20 es relativamente sencilla y se resume en la tabla 8.21. (Siempre se indicará la variable básica entrante al colocar un signo + dentro de su celda y al dejar el valor correspondiente de $c_{ij} - u_i - v_j$ en la esquina inferior derecha de esa celda.) Al aumentar x_{25} debe disminuir x_{15} en la misma cantidad para conservar la demanda de 60 en la columna 5; esto a su vez requiere que se aumente x_{13} en esta cantidad para mantener la oferta de 50 en el renglón 1, lo cual, de nueva cuenta, exige una disminución del valor de x_{23} para conservar la demanda de 70 en la columna 3. Esta disminución de x_{23} completa con éxito la reacción en cadena, ya que también conserva la oferta del renglón 2. (De manera equivalente, se pudo haber iniciado la reacción en cadena para esta cantidad de recursos en el renglón 2 con una disminución de x_{23}, y después la reacción en cadena continuaría con un aumento de x_{13} y una disminución de x_{15}.)

El resultado final es que las celdas (2, 5) y (1, 3) se convierten en **celdas receptoras**, cada una con su asignación adicional proveniente de las **celdas donadoras** (1, 5) y (2, 3.) (Estas celdas están indicadas en la tabla 8.21 por medio de los signos + y −.) Observe que tuvo que elegirse la celda (1, 5) como donadora para la columna 5 y no la (4, 5), puesto que esta última no hubiera tenido celda receptora en el renglón 4 para continuar la reacción en cadena. [De manera similar,

■ **TABLA 8.21** **Parte inicial de la tabla símplex de transporte que muestra la reacción en cadena que se produce al incrementar la variable básica entrante x_{25}**

			Destino 3	4	5	**Recursos**
Origen	1	…	[13] (40)+	[22] +4	[17] (10)−	50
	2	…	[13] (30)−	[19] +1	[15] [+] −2	60
		…	…	…	…	
Demanda			70	30	60	

si la reacción en cadena se iniciara en el renglón 2, la celda (2, 1) no podría ser donadora, pues no sería posible completar la reacción en cadena después de elegir, sin otra opción, la celda (3, 1), ya que ni la celda (3, 2) ni la (3, 4), como donadoras, tienen celda receptora.] Observe, además, que, a excepción de la variable básica entrante, *todas* las celdas receptoras y donadoras de la reacción en cadena deben corresponder a variables *básicas* en la solución BF actual.

Cada celda donadora disminuye su asignación en una cantidad exactamente igual al aumento que tiene la variable básica entrante (y las otras celdas receptoras). Entonces, la celda donadora que comienza con la asignación más pequeña —en este caso la celda (1, 5), puesto que $10 < 30$ en la tabla 8.21— debe ser la primera en llegar a una asignación de cero conforme se incrementa la variable entrante x_{25}. Así, x_{15} se convierte en la variable básica saliente.

En general, siempre existe sólo *una* reacción en cadena —en cualquier dirección— que se puede completar con éxito para conservar la factibilidad, cuando la variable básica entrante aumenta su valor. Esta reacción en cadena se identifica al hacer una selección entre las celdas que tienen variables básicas: primero, la celda donadora de la *columna* con la variable básica; después, la celda receptora en el renglón que corresponde a la celda donadora; luego, la celda donadora de la columna donde se encuentra esta celda receptora, y así sucesivamente, hasta que la reacción en cadena conduce a una celda donadora en el *renglón* con la variable básica entrante. Cuando una columna o renglón tiene más de una celda adicional con variable básica, puede ser necesario explorar el camino a seguir para averiguar cuál debe seleccionarse como donadora o receptora. (Todas las demás, menos la adecuada, llegarán en algún momento a un camino sin salida en un renglón o columna sin otra celda con variable básica.) Después de identificar la reacción en cadena, la celda donadora que tiene la asignación *menor* proporciona en forma automática la variable básica saliente. (En caso de un empate para la celda donadora, se puede elegir cualquiera como variable básica saliente.)

Paso 3: Encontrar la nueva solución BF. La *nueva solución BF* se identifica mediante la suma del valor (antes de los cambios) de la variable básica saliente a las asignaciones de cada celda receptora y la resta de *esta misma cantidad* de las asignaciones de cada celda donadora. En la tabla 8.21 se observa que el valor de la variable básica que sale x_{15} es 10, por lo que esta porción de la tabla símplex de transporte cambia, como se ilustra en la tabla 8.22 para la nueva solución. (Como x_{15} es no básica en la nueva solución, su nueva asignación es cero y ya no se muestra en esta tabla.)

Ahora se puede señalar una interpretación útil de las cantidades $c_{ij} - u_i - v_j$ que se obtuvieron en la prueba de optimalidad. Debido al cambio de 10 unidades en las asignaciones de las celdas donadoras a las receptoras —que se muestra en las tablas 8.21 y 8.22—, el costo total cambia en

$$\Delta Z = 10(15 - 17 + 13 - 13) = 10(-2) = 10(c_{25} - u_2 - v_5).$$

Así, el efecto de aumentar el valor de la variable básica entrante x_{25} a partir de cero fue un cambio en el costo a una tasa de -2 por unidad de aumento. Esto es justo lo que indica el valor de $c_{25} - u_2 - v_5 = -2$ en la tabla 8.20. De hecho, otra manera (menos eficiente) de obtener $c_{ij} - u_i - v_j$ para cada variable no básica x_{ij} es identificar la reacción en cadena que causaría un aumento de una

■ **TABLA 8.22 Parte de la segunda tabla símplex de transporte que muestra los cambios en la solución BF**

		Destino				
			3	4	5	Recursos
Origen	1	…	13 (50)	22	17	50
	2	…	13 (20)	19	15 (10)	60
		…	…	…	…	
Demanda			70	30	60	

unidad en esta variable y calcular el cambio en el costo. Algunas veces es útil esta interpretación intuitiva para verificar los cálculos en la prueba de optimalidad.

Antes de completar la solución del problema del Metro Water District, se hará un resumen de las reglas del método símplex de transporte.

Resumen del método símplex de transporte

Paso inicial: Se construye una solución BF con el procedimiento descrito antes en esta sección. Se realiza la prueba de optimalidad

Prueba de optimalidad: Se obtiene u_i y v_j al elegir el renglón con el mayor número de asignaciones y establecer su $u_i = 0$; después se resuelve el sistema de ecuaciones $c_{ij} = u_i + v_j$ para cada (i, j) tal que x_{ij} es básica. Si $c_{ij} - u_i - v_j \geq 0$ para toda (i, j) tal que x_{ij} es *no básica*, entonces la solución actual es óptima, por lo que el proceso se detiene. De lo contrario, se realiza una iteración.

Iteración:

1. Se determina la variable básica entrante: se elige la variable no básica x_{ij} que tiene el valor *más negativo* —en términos absolutos— para $c_{ij} - u_i - v_j$.

2. Se determina la variable básica saliente: se identifica la reacción en cadena que se necesita para conservar la factibilidad cuando aumenta el valor de la variable básica entrante. Entre las celdas donadoras se selecciona la variable básica que tiene el *menor* valor.

3. Se determina la nueva solución BF: se suma el valor de la variable básica saliente a las asignaciones de las celdas receptoras y se resta a las asignaciones de las celdas donadoras.

Para continuar con la aplicación de este procedimiento al problema del Metro Water District, se llega al conjunto completo de tablas símplex de transporte que se muestra en la tabla 8.23. Como en la cuarta tabla todas las $c_{ij} - u_i - v_j$ son no negativas, la prueba de optimalidad identifica este conjunto de asignaciones como óptimo, y concluye el algoritmo.

Para el lector, una buena práctica sería obtener los valores de las u_i y v_j que se dan en la segunda, tercera y cuarta iteraciones. Inténtelo mediante la manipulación directa de la tabla. También verifique las reacciones en cadena de la segunda y tercera iteraciones, que son un poco más complicadas que la que se muestra en la tabla 8.21.

Características especiales de este ejemplo

Se deben observar tres puntos importantes ilustrados en este ejemplo. Primero, la solución inicial básica factible es *degenerada*, puesto que la variable básica $x_{31} = 0$. Pero esto no causa complicaciones, porque la celda (3, 1) se convierte en *celda receptora* en la segunda tabla, lo que hace que x_{31} aumente a un valor mayor que cero.

Segundo, en la tercera tabla surge otra variable básica degenerada (x_{34}), puesto que dos variables básicas en *dos* celdas donadoras de la segunda tabla, la (2, 1) y la (3, 4) están *empatadas* pues tienen el mismo valor más pequeño (30). (Este empate se rompe de manera arbitraria mediante la elección de x_{21} como la variable básica que sale; si se hubiera seleccionado x_{34} entonces x_{21} se habría convertido en la variable básica degenerada.) Parece que esta variable básica degenerada crea algunas complicaciones subsecuentes, porque la celda (3, 4) se convierte en *celda donadora* en la tercera iteración, pero ¡no tiene nada que donar! Por fortuna, no hay que preocuparse. Como la cantidad que se va a sumar y restar en las celdas receptoras y donadoras es cero, las asignaciones no cambian. Sin embargo, la variable básica degenerada sí se convierte en la variable básica saliente, y se sustituye por la variable básica entrante con una asignación de cero, como se muestra en la cuarta tabla. Este cambio en el conjunto de variables básicas modifica los valores de u_i y v_j. Si cualquiera de las $c_{ij} - u_i - v_j$ hubiera sido negativa en la cuarta tabla, el algoritmo hubiera seguido adelante para hacer cambios reales en las asignaciones, siempre y cuando todas las celdas donadoras tuvieran variables básicas no degeneradas.

Tercero, como ninguna de las $c_{ij} - u_i - v_j$ resultó negativa en la cuarta tabla, el conjunto equivalente de asignaciones de la tercera iteración también es óptimo. Así, el algoritmo realizó una iteración más de las necesarias. Esta iteración adicional es un defecto que surge de manera ocasional tanto en el método símplex como en el de transporte debido a la degeneración, pero no es tan grave como para justificar algún ajuste a los algoritmos.

■ **TABLA 8.23 Conjunto completo de tablas símplex de transporte del problema del Metro Water District**

Iteración 0		**Destino** 1	2	3	4	5	**Recursos**	u_i
Origen	1	[16] +2	[16] +2	[13] (40)+	[22] +4	[17] (10)−	50	−5
	2	[14] (30)	[14] 0	[13] (30)−	[19] +1	[15] [+] −2	60	−5
	3	[19] (0)	[19] (20)	[20] +2	[23] (30)	[M] $M - 22$	50	0
	4(D)	[M] $M + 3$	[0] +3	[M] $M + 4$	[0] −1	[0] (50)	50	−22
Demanda		30	20	70	30	60	Z = 2 570	
	v_j	19	19	18	23	22		

Iteración 1		**Destino** 1	2	3	4	5	**Recursos**	u_i
Origen	1	[16] +2	[16] +2	[13] (50)	[22] +4	[17] +2	50	−5
	2	[14] (30)−	[14] 0	[13] (20)	[19] +1	[15] (10)+	60	−5
	3	[19] (0)+	[19] (20)	[20] +2	[23] (30)−	[M] $M - 20$	50	0
	4(D)	[M] $M + 1$	[0] +1	[M] $M + 2$	[0] [+] −3	[0] (50)−	50	−20
Demanda		30	20	70	30	60	Z = 2 550	
	v_j	19	19	18	23	20		

Iteración 2		**Destino** 1	2	3	4	5	**Recursos**	u_i
Origen	1	[16] +5	[16] +5	[13] (50)	[22] +7	[17] +2	50	−8
	2	[14] +3	[14] +3	[13] (20)−	[19] +4	[15] (40)+	60	−8
	3	[19] (30)	[19] (20)	[20] [+] −1	[23] (0)−	[M] $M - 23$	50	0
	4(D)	[M] $M + 4$	[0] +4	[M] $M + 2$	[0] (30)+	[0] (20)−	50	−23
Demanda		30	20	70	30	60	Z = 2 460	
	v_j	19	19	21	23	23		

TABLA 8.23 (*continuación*)

Iteración 3		Destino 1	Destino 2	Destino 3	Destino 4	Destino 5	Recurso	u_i
Origen	1	16; +4	16; +4	13; (50)	22; +7	17; +2	50	−7
	2	14; +2	14; +2	13; (20)	19; +4	15; (40)	60	−7
	3	19; (30)	19; (20)	20; (0)	23; +1	M; $M - 22$	50	0
	4(*D*)	M; $M + 3$	0; +3	M; $M + 2$	0; (30)	0; (20)	50	−22
Demanda		30	20	70	30	60	$Z = 2\,460$	
	v_j	19	19	20	22	22		

Si se desea consultar ejemplos adicionales (más pequeños) de la aplicación del método símplex de transporte, hay dos más disponibles. Uno es la demostración proporcionada por el área de problemas de transporte del OR Tutor. El **otro ejemplo** puede encontrarse en la sección Worked Examples del sitio de internet de este libro, donde se incluye un ejemplo de este tipo. También, en el IOR Tutorial se pueden encontrar tanto una rutina interactiva, como una automática, para este método.

Ahora que se ha estudiado el método símplex de transporte, el lector se encuentra en posición de verificar por sí mismo la manera en que el algoritmo proporciona, de hecho, una demostración de la *propiedad de soluciones enteras* que se presentó en la sección 8.1. El problema 8.2-20 le ayuda a guiarlo a través del razonamiento.

8.3 PROBLEMA DE ASIGNACIÓN

El **problema de asignación** es un tipo especial de problema de programación lineal en el que los **asignados** son recursos que se destinan a la realización de **tareas**. Por ejemplo, los asignados pueden ser empleados a quienes se tiene que dar trabajo. La asignación de personas a trabajos es una aplicación común del problema de asignación.[10] Sin embargo, los asignados no tienen que ser personas. También pueden ser máquinas, vehículos o plantas, o incluso periodos a los que se asignan tareas. El primero de los siguientes ejemplos se refiere a máquinas asignadas a lugares, de manera que la tarea en este caso se trata sólo de tener una máquina. Un ejemplo posterior se refiere a plantas a las que se asignan productos que deben fabricar.

Para que se ajuste a la definición de un problema de asignación, es necesario que este tipo de aplicaciones se formule de manera tal que se cumplan los siguientes supuestos.

1. El número de asignados es igual al número de tareas. (Este número se denota por n.)
2. A cada asignado se le asigna sólo *una* tarea.
3. Cada tarea debe realizarla sólo *un* asignado.
4. Existe un costo c_{ij} asociado con el asignado i ($i = 1, 2, \ldots, n$) que realiza la tarea j ($j = 1, 2, \ldots, n$).
5. El objetivo es determinar cómo deben hacerse las n asignaciones para minimizar los costos totales.

[10] Por ejemplo, vea L. J. LeBlanc, D. Randels, Jr. y T. K. Swann, "Heery International's Spreadsheet Optimization Model for Assigning Managers to Construction Projects", en *Interfaces*, **30**(6): 95-106, noviembre-diciembre de 2000. También, en la página 98 de este artículo se mencionan siete aplicaciones distintas del problema de asignación.

Cualquier problema que satisface todos estos supuestos se puede resolver en forma muy eficiente mediante los algoritmos diseñados de manera especial para los problemas de asignación.

Los primeros tres supuestos son bastante restrictivos. Muchas aplicaciones potenciales no las satisfacen por completo. Con frecuencia es posible reformular el problema para hacerlo que se ajuste. Por ejemplo, muchas veces se pueden usar *asignados ficticios* o *tareas ficticias* con este fin. En los ejemplos se ilustran estas técnicas de formulación.

Ejemplo prototípico

La JOB SHOP COMPANY compró tres máquinas nuevas de diferentes tipos. Existen cuatro sitios disponibles dentro del taller en donde se podría instalar una máquina. Algunos de ellos son más adecuados que otros para ciertas máquinas en particular por su cercanía a los centros de trabajo que tendrían un flujo intenso de trabajo hacia y desde estas máquinas. (No habrá flujo de trabajos *entre* las nuevas máquinas.) Por tanto, el objetivo es asignar las nuevas máquinas a los lugares disponibles de manera que se minimice el costo total del manejo de materiales. En la tabla 8.24 se proporciona el costo estimado por unidad de tiempo del manejo de los materiales en cuestión, con cada una de las máquinas en los sitios respectivos. El lugar 2 no se considera adecuado para la máquina 2 por lo que no se proporciona un costo para este caso.

Para formular éste como un problema de asignación, debe introducirse una *máquina ficticia* en el lugar adicional. Además, debe asignarse un costo muy grande M a la asignación de la máquina 2 en el lugar 2 para evitarla en la solución óptima. En la tabla 8.25 se muestra la tabla de costos que resulta para este problema de asignación. Esta *tabla de costos* contiene todos los datos necesarios para resolver el problema. La solución óptima es asignar la máquina 1 al lugar 4, la máquina 2 al lugar 3 y la máquina 3 al lugar 1 con un costo total de $29 por hora. La máquina ficticia se asigna al lugar 2, con lo que esta posición queda disponible para alguna asignación de máquina real futura.

Más adelante se discutirá cómo se obtiene esta solución después de formular el modelo matemático para el problema general de asignación.

Modelo del problema de asignación

El modelo matemático para manejar el problema de asignación utiliza las siguientes variables de decisión:

$$x_{ij} = \begin{cases} 1 & \text{si se asigna } i \text{ para realizar la tarea } j, \\ 0 & \text{si no es así,} \end{cases}$$

para $i = 1, 2, \ldots, n$ y $j = 1, 2, \ldots, n$. Entonces, cada x_{ij} es una *variable binaria* (toma valores 0 o 1). Las variables binarias son importantes en investigación de operaciones para representar las *decisiones de sí* o *no*, como se verá en detalle en el capítulo de programación entera (capítulo 11). En este caso, las *decisiones de sí* o *no* son: ¿debe el asignado i realizar la tarea j?

Si Z es el costo total, el modelo del problema de asignación es

$$\text{Minimizar} \qquad Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij},$$

TABLA 8.24 Costos de manejo de materiales de la Job Shop Co. ($)

		Localidad			
		1	2	3	4
	1	13	16	12	11
Máquina	2	15	—	13	20
	3	5	7	10	6

TABLA 8.25 Tabla de costos del problema de asignación de la Job Shop Co.

		Tarea (localidad)			
		1	2	3	4
	1	13	16	12	11
Asignado	2	15	M	13	20
(máquina)	3	5	7	10	6
	4(D)	0	0	0	0

sujeta a

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad \text{para } i = 1, 2, \ldots, n,$$

$$\sum_{i=1}^{n} x_{ij} = 1 \qquad \text{para } j = 1, 2, \ldots, n,$$

y

$$x_{ij} \geq 0, \qquad \text{para toda } i \text{ y } j$$
$$(x_{ij} \text{ binaria}, \qquad \text{para toda } i \text{ y } j).$$

El primer conjunto de restricciones funcionales especifica que cada asignado realice sólo una asignación, mientras que el segundo conjunto requiere que cada asignación sea realizada sólo por un asignado. Si se elimina la restricción entre paréntesis de que x_{ij} sean binarias, resulta claro que el modelo es un tipo especial de problema de programación lineal, por lo que se puede resolver de inmediato. Por fortuna, debido a las razones que se expondrán en seguida, se *puede* eliminar esta restricción. (Esta eliminación es la razón por la que el problema de asignación se incluye en este capítulo y no en el de programación entera.)

Ahora se hará una comparación de este modelo —sin la restricción binaria— con el modelo del problema de transporte que se presentó en la segunda subsección de la sección 8.1 (incluso la tabla 8.6). Observe que sus estructuras son similares. En realidad, el problema de asignación es sólo un caso especial de los problemas de transporte en donde los *orígenes* son ahora los *asignados* y los *destinos* son las *asignaciones* o *tareas* y donde

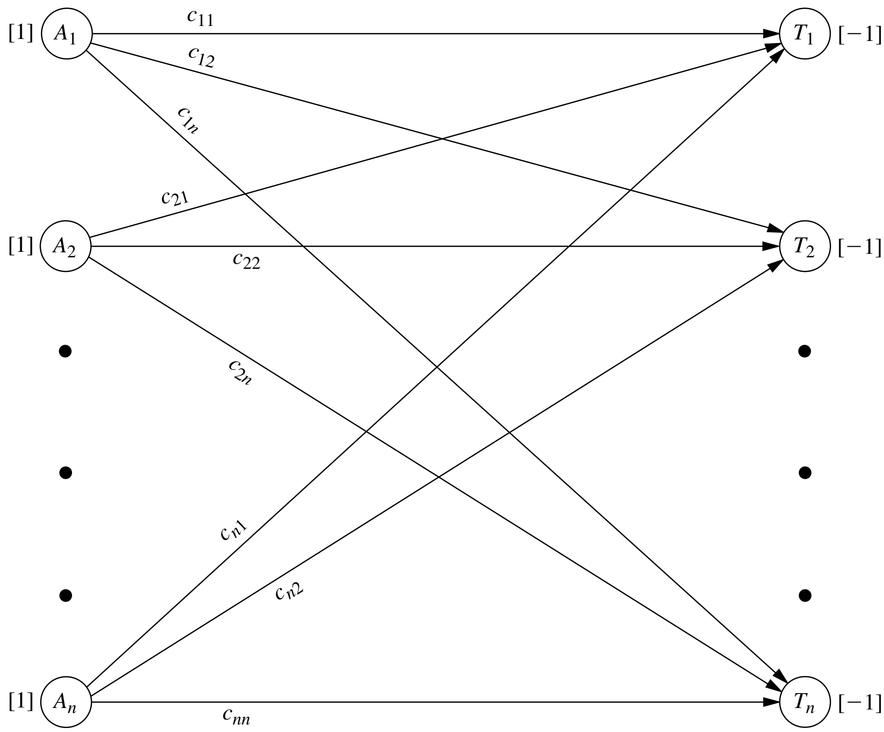
Número de orígenes m = número de destinos n,
Cada recurso $s_i = 1$,
Cada demanda $d_j = 1$.

Ahora se enfocará la atención en la **propiedad de soluciones enteras** de la subsección sobre el modelo de transporte. Como ahora toda s_i y d_j son enteros (= 1), esta propiedad significa que *toda solución BF* (incluso la óptima) es una solución *entera* para un problema de asignación. Las restricciones funcionales del modelo de asignación evitan que las variables sean mayores que 1, y las restricciones de no negatividad impiden que existan valores menores que cero. Por tanto, si se elimina la restricción binaria para poder resolver el problema de asignación como un problema de programación lineal, las soluciones BF que se obtienen (incluso la solución óptima final) satisfarán en forma *automática* la restricción binaria.

Igual que el problema de transporte se puede representar con una red (vea la figura 8.3), el problema de asignación puede describirse de manera similar, como lo muestra la figura 8.5. En este caso la primera columna enumera los n asignados y la segunda las n tareas. Los números entre corchetes indican el número de asignados que se proporcionan en ese lugar de la red, por lo cual los valores de la izquierda son 1 de manera automática, mientras que los valores de -1 de la derecha indican que cada tarea utiliza un asignado.

Por lo general, los que aplican esta técnica a problemas de asignación específicos no se toman la molestia de escribir todo el modelo matemático. Es mucho más sencillo formularlo en una tabla de costos (como la 8.25), con la identificación de los asignados y las asignaciones, puesto que contiene todos los datos esenciales en una forma mucho más compacta.

Algunas veces surgen problemas que no se ajustan del todo al modelo de un problema de asignación, porque ciertos asignados realizarán más de una tarea. En este caso, el problema se puede reformular para ajustarse al modelo al dividir cada uno de esos asignados en asignados nuevos separados (pero idénticos) donde cada nuevo asignado realizará sólo una tarea. (En la tabla 8.29 se ilustrará este asunto para un ejemplo subsecuente.) De manera similar, si una tarea será realizada por asignados múltiples, ésta se puede dividir en tareas nuevas separadas (pero idénticas) donde cada nueva tarea será desarrollada sólo por un asignado de acuerdo con el modelo reformulado. En la sección Worked Examples en el sitio de internet de este libro se proporciona un ejemplo que ilustra ambos casos y la reformulación para ajustarse al modelo de un problema de asignación. También se muestra una formulación alternativa como un problema de transporte.



■ **FIGURA 8.5**
Representación de red del problema de asignación.

Procedimientos de solución de problemas de asignación

Se dispone de procedimientos alternativos de solución para resolver problemas de asignación. Aquellos que no son mucho más grandes que el de la Job Shop Co., se pueden resolver con rapidez mediante el método símplex general, y quizá sea conveniente usar un paquete básico —como Excel y su Solver— que utilice este método. Si se aplicara este enfoque al problema de la Job Shop Co., no habría sido necesario agregar la máquina ficticia en la tabla 8.25 para ajustar el modelo de asignación. Las restricciones sobre el número de máquinas asignadas a cada lugar se expresarían como

$$\sum_{i=1}^{3} x_{ij} \leq 1 \qquad \text{para } j = 1, 2, 3, 4.$$

Como se muestra en el archivo de Excel de este capítulo, una formulación en hoja de cálculo sería muy similar a la del problema de transporte que se describió en la figura 8.4, excepto que ahora todos los suministros y demandas serían 1 y las restricciones de demanda serían del tipo ≤ 1 en lugar de $= 1$.

Sin embargo, los problemas de asignación grandes se pueden resolver mucho más rápido si se usan procedimientos de solución más específicos, por lo que en ese caso se recomienda usarlos en lugar del método símplex general.

Debido a que el problema de asignación es un tipo especial de problema de transporte, una forma conveniente de resolverlo es aplicar el método símplex de transporte que se describió en la sección 8.2. Este enfoque requiere que se convierta la tabla de costos en una tabla de parámetros para el problema de transporte equivalente, como se muestra en la tabla 8.26*a*.

Por ejemplo, la tabla 8.26*b* muestra la tabla de parámetros del problema de la Job Shop Co., que se obtuvo a partir de la tabla de costos del cuadro 8.25. Cuando se aplica el método símplex de transporte a esta formulación, la solución óptima tiene las variables básicas $x_{13} = 0$, $x_{14} = 1$, $x_{23} = 1$, $x_{31} = 1$, $x_{41} = 0$, $x_{42} = 1$, $x_{43} = 0$. (En el problema 8.3-6 se pide que se verifique esta solución.) Las variables básicas degeneradas ($x_{ij} = 0$) y la asignación de la máquina ficticia ($x_{42} = 1$) no tienen significado para el problema original, por lo cual las asignaciones reales son la máquina 1 al lugar 4, la 2 al lugar 3 y la 3 al lugar 1.

TABLA 8.26 Tabla de parámetros del problema de asignación formulado como un problema de transporte, ilustrado para el ejemplo de la Job Shop Co.

***a*) Caso general**

		Costo por unidad distribuida			
		Destino			
	1	2	⋯	n	Recursos
Origen 1	c_{11}	c_{12}	⋯	c_{1n}	1
2	c_{21}	c_{22}	⋯	c_{2n}	1
⋮	⋯	⋯	⋯	⋯	⋮
$m = n$	c_{n1}	c_{n2}	⋯	c_{nn}	1
Demanda	1	1	⋯	1	

***b*) Ejemplo de la Job Shop Co.**

		Costo por unidad distribuida			
		Destino (localidad)			
	1	2	3	4	Recursos
Origen (máquina) 1	13	16	12	11	1
2	15	M	13	20	1
3	5	7	10	6	1
4(*D*)	0	0	0	0	1
Demanda	1	1	1	1	

No es coincidencia que esta solución óptima proporcionada por el método símplex de transporte tenga tantas variables básicas degeneradas. En el caso de cualquier problema de asignación en el que se tienen que hacer n asignaciones, la formulación del problema de transporte que se muestra en la tabla 8.26*a* tiene $m = n$, es decir, el número de orígenes (m) y el número de destinos (n) son iguales al número de asignaciones (n). En general, los problemas de transporte tienen $m + n - 1$ variables básicas (asignaciones); de esta forma, toda solución BF de este tipo de problema de transporte tiene $2n - 1$ variables básicas, pero exactamente n de estas x_{ij} son iguales a 1, correspondientes a las n asignaciones. Por tanto, como todas las variables son binarias, siempre habrá $n - 1$ variables básicas degeneradas ($x_{ij} = 0$). Se dijo al final de la sección 8.2 que las variables básicas degeneradas no causan mayor complicación para ejecutar el algoritmo. No obstante, a menudo ocasionan *iteraciones en exceso*, en las que nada cambia —las mismas asignaciones— excepto la etiqueta de qué asignaciones de cero corresponden a variables básicas degeneradas en lugar de hacerlo a variables no básicas. Estas iteraciones desperdiciadas son un inconveniente notable cuando se desea aplicar el método símplex de transporte a estas situaciones en las que *siempre* hay tantas variables básicas degeneradas.

Otro inconveniente del método símplex de transporte es que se trata sólo de un algoritmo *con fines generales* para resolver todos los problemas de transporte. Por esto, no aprovecha la estructura especial adicional en este tipo especial de problemas ($m = n$, toda $s_i = 1$, y toda $d_j = 1$). Por fortuna, se han desarrollado algoritmos especiales que simplifican el procedimiento para la solución exclusiva de los problemas de asignación. Estos algoritmos operan de manera directa sobre la tabla de costos y no toman en cuenta las variables básicas degeneradas. Si se dispone de un paquete de computadora para alguno de estos algoritmos, será preferible usarlo en lugar del método símplex de transporte, en especial en problemas muy grandes.[11]

En la sección 8.4 se describe uno de estos algoritmos especializados —llamado *algoritmo húngaro*— para resolver sólo problemas de asignación de una manera muy eficiente.

En el IOR Tutorial se incluye tanto una rutina interactiva como una automática para aplicar este algoritmo.

Ejemplo: asignación de productos a plantas

La BETTER PRODUCTS COMPANY ha decidido iniciar la fabricación de cuatro nuevos productos en tres plantas que por el momento tienen exceso de capacidad de producción. Los productos requieren de un esfuerzo productivo comparable por unidad, por lo que la capacidad de producción disponible en las plantas se mide por el número de unidades de cualquier producto que se pueden fabricar por día, como se muestra en la última columna de la tabla 8.27. El último renglón muestra

[11] Un artículo que compara varios algoritmos del problema de asignación se encuentra en J. L. Kennington y Z. Wang, "An Empirical Analysis of the Dense Assignment Problem: Sequential and Parallel Implementations", en *ORSA Journal on Computing*, **3**: 299-306, 1991.

■ **TABLA 8.27** Datos del problema de la Better Products Co.

		Costo unitario por producto ($)				Capacidad disponible
		1	2	3	4	
	1	41	27	28	24	75
Planta	2	40	29	—	23	75
	3	37	30	27	21	45
Tasa de producción		20	30	30	40	

la producción diaria que se requiere para satisfacer las ventas proyectadas. Cada planta puede producir cualquiera de estos productos, *excepto* la planta 2 que *no* puede fabricar el producto 3. Sin embargo, el costo variable por unidad de cada producto difiere entre una planta y otra, como se muestra en el cuerpo principal de la tabla 8.27.

La administración necesita tomar la decisión de cómo dividir la producción entre las plantas. Tiene dos opciones:

Opción 1: Permitir la *separación de productos*, de tal manera que el mismo producto se pueda fabricar en más de una planta.
Opción 2: No autorizar la *separación de productos*.

La segunda opción impone una restricción que sólo puede aumentar el costo de una solución óptima según la tabla 8.27. Por otro lado, la ventaja clave de la opción 2 es que elimina algunos *costos ocultos* asociados con la separación de productos que no se reflejan en la tabla 8.27, que incluyen costos adicionales de preparación, distribución y administración. Por ello, la gerencia quiere que se analicen ambas opciones antes de tomar la decisión final. Además, en el caso de la opción 2 se ha especificado que debe asignarse al menos uno de los productos a cada planta.

Se formulará y resolverá el modelo de acuerdo con cada opción, donde la opción 1 conduce al problema de transporte y la opción 2 al problema de asignación.

Formulación de la opción 1. Al permitir la separación de productos, la tabla 8.27 se puede convertir en forma directa en una tabla de parámetros para un problema de transporte. Las plantas se convierten en orígenes y los productos en destinos (o viceversa); de este modo, los recursos se interpretan como las capacidades de producción y las demandas como las tasas de producción que se requieren. Sólo se tienen que hacer dos cambios en la tabla 8.27. Primero, como la planta 2 no puede fabricar el producto 3, se evita esa asignación al darle un costo unitario muy grande, M. Segundo, la capacidad total ($75 + 75 + 45 = 195$) excede la producción total que se requiere ($20 + 30 + 30 + 40 = 120$), por lo cual se necesita un destino ficticio con una demanda de 75 para balancear estas dos cantidades. La tabla de parámetros que resulta se muestra en la tabla 8.28.

■ **TABLA 8.28** Tabla de parámetros para formular el problema de transporte de la opción 1 del problema de la Better Products Co.

		Costo por unidad distribuida					
		Destino (producto)					
		1	2	3	4	5(*D*)	Recurso
Origen (planta)	1	41	27	28	24	0	75
	2	40	29	*M*	23	0	75
	3	37	30	27	21	0	45
Demanda		20	30	30	40	75	

La solución óptima de este problema de transporte tiene variables básicas (asignaciones) $x_{12} = 30$, $x_{13} = 30$, $x_{15} = 15$, $x_{24} = 15$, $x_{25} = 60$, $x_{31} = 20$ y $x_{34} = 25$, de manera que:

La planta 1 produce todo de los productos 2 y 3.
La planta 2 produce 37.5% del producto 4.
La planta 3 produce 62.5% del producto 4 y todo del producto 1.

El costo total es $Z = \$3\,260$ diarios.

Formulación de la opción 2. Si no se separan los productos, cada uno de éstos debe asignarse a una sola planta. Entonces, la fabricación de productos se puede interpretar como las tareas en un problema de asignación y las plantas como los asignados.

La administración ha especificado que debe asignarse al menos uno de los productos a cada planta. Se debe elaborar más productos (cuatro) que plantas (tres), por lo que se tendrá que asignar dos productos a una de las plantas. La planta 3 tiene apenas la capacidad adicional para fabricar un producto (vea la tabla 8.27), por lo que la planta 1 o *bien* la 2 fabricará el otro producto.

Para hacer posible la asignación de este producto adicional dentro de la formulación de un problema de asignación, las plantas 1 y 2 se dividen en dos asignados cada una, como se muestra en la tabla 8.29.

El número de asignados (ahora cinco) debe ser igual al número de tareas (ahora cuatro), así que se introduce una *asignación ficticia* (producto) como 5(D) en la tabla 8.29. El papel de esta asignación ficticia es proporcionar un segundo producto ficticio a cualquiera de las plantas 1 o 2, que recibe sólo un producto real. No se incurre en costos por fabricar un producto ficticio así que, como siempre, los costos de la tarea ficticia son cero. La excepción es el costo M del último renglón de la tabla 8.29. La razón es que se debe asignar un producto real a la planta 3 —elegido entre los productos 1, 2, 3 o 4— y se necesita el método de la gran M para evitar que se asigne el producto ficticio a esta planta. (Igual que en la tabla 8.28, M se usa con el fin de evitar la asignación no factible del producto 3 a la planta 2.)

El resto de los costos de la tabla 8.29 no son los costos unitarios que se muestran en la tabla 8.27 u 8.28. La tabla 8.28 proporciona una formulación para el problema de transporte —en el caso de la opción 1—, de manera que los costos unitarios son apropiados en ese caso, pero ahora se quiere formular un problema de asignación, en el caso de la opción 2. Para un problema de asignación, el costo c_{ij} es el costo *total* asociado al hecho de que el asignado i realiza la tarea j. En la tabla 8.29, el *costo total* (por día) para que la planta i fabrique el producto j es el costo de producción multiplicado por el número de unidades producidas (por día), donde estas dos cantidades que se deben multiplicar se dan por separado en la tabla 8.27. Por ejemplo, considere la asignación de la planta 1 al producto 1. Si se usa el costo unitario de la tabla 8.28 ($41) y la demanda —número de unidades producidas por día— de la misma tabla (20), se obtiene

Costo de fabricar una unidad del producto 1 en la planta 1	= \$41
Producción (diaria) requerida del producto 1	= 20 unidades
Costo total (diario) de asignar la planta 1 al producto 1	= 20 (\$41)
	= \$820

TABLA 8.29 Tabla de costos para formular el problema de asignación de la opción 2 del problema de la Better Products Co.

		Tarea (producto)				
		1	**2**	**3**	**4**	**5(*D*)**
Asignado (planta)	1*a*	820	810	840	960	0
	1*b*	820	810	840	960	0
	2*a*	800	870	*M*	920	0
	2*b*	800	870	*M*	920	0
	3	740	900	810	840	*M*

de forma que 820 es el elemento de la tabla 8.29 para el costo del asignado 1*a* o 1*b* si realiza la tarea 1.

La solución óptima de este problema de asignación es la siguiente:

La planta 1 fabrica los productos 2 y 3.
La planta 2 fabrica el producto 1.
La planta 3 fabrica el producto 4.

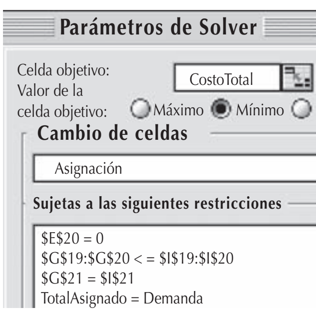
En este caso, la asignación ficticia se da a la planta 2. El costo total es $Z = \$3\,290$ por día.

Como es usual, una manera de obtener esta solución óptima es convertir la tabla de costos 8.29 en una tabla de parámetros equivalente para el problema de transporte (vea la tabla 8.26) y después aplicar el método símplex de transporte. Debido a que la tabla 8.29 contiene dos renglones idénticos, este enfoque se puede simplificar combinando los cinco asignados en tres orígenes con costos 2, 2 y 1, respectivamente. (Vea el problema 8.3-5.) Esta simplificación también reducirá el número de variables básicas degeneradas en todas las soluciones BF. Por tanto, aunque esta formulación simplificada ya no se ajusta al formato que se presentó en la tabla 8.26*a* para un problema de asignación, es una formulación más eficiente para aplicar el método símplex de transporte.

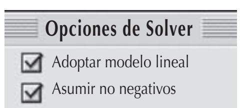
En la figura 8.6 se muestra cómo se puede usar Excel y su herramienta Solver para obtener esta solución óptima, que se despliega en las celdas cambiantes Asignación(C19:F21) de la hoja de cálculo. Como se usa el método símplex general, no hay necesidad de ajustar la formulación al

FIGURA 8.6
Formulación de la opción 2 en hoja de cálculo para el problema de la Better Products Co., como una variante del problema de asignación. La celda objetivo es CostoTotal (I24) y las otras celdas de salida son Costo (C12:F14), AsignacionesTotales (G19:G21) y TotalAsignado (C22:F22), donde las ecuaciones introducidas en estas celdas se muestran debajo de la hoja de cálculo. Los valores de 1 en las celdas cambiantes de asignación (C19:F21) presentan el plan de producción óptimo obtenido por el Solver.

	A	B	C	D	E	F	G	H	I
1	**Better Products Co. Production Planning Problem (Revised)**								
2									
3		**Unit Cost**	Product 1	Product 2	Product 3	Product 4			
4		Plant 1	$41	$27	$28	$24			
5		Plant 2	$40	$29	-	$23			
6		Plant 3	$37	$30	$27	$21			
7									
8		Required Production	20	30	30	40			
9									
10									
11		**Cost ($/day)**	Product 1	Product 2	Product 3	Product 4			
12		Plant 1	$820	$810	$840	$960			
13		Plant 2	$800	$870	-	$920			
14		Plant 3	$740	$900	$810	$840			
15									
16									
17							Total		
18		**Assignment**	Product 1	Product 2	Product 3	Product 4	Assignments		Supply
19		Plant 1	0	1	1	0	2		2
20		Plant 2	1	0	0	0	1		2
21		Plant 3	0	0	0	1	1	=	1
22		Total Assigned	1	1	1	1			
23			=	=	=	=			Total Cost
24		Demand	1	1	1	1			$ 3,290



	B	C	D	E	F
11	**Cost ($/day)**	Product 1	Product 2	Product 3	Product 4
12	Plant 1	=C4*C8	=D4*D8	=E4*E8	=F4*F8
13	Plant 2	=C8*C5	=D8*D5	-	=F8*F5
14	Plant 3	=C8*C6	=D8*D6	=E8*E6	=F8*F6



	G
17	Total
18	Asignados
19	=SUMA(C19:F19)
20	=SUMA(C20:F20)
21	=SUMA(C21:F21)

	B	C	D	E	F
22	Total Asignados	=SUMA(C19:C21)	=SUMA(D19:D21)	=SUMA(E19:E21)	=SUMA(F19:F21)

Nombre de rango	Celdas
Asignación	C19:F21
Costo	C12:F14
Demanda	C24:F24
ProducciónRequerida	C8:F8
Suministro	I19:I21
TotalAsignado	C22:F22
TotalDeAsignaciones	G19:G21
CostoTotal	I24
CostoUnitario	C4:F6

	I
23	Costo total
24	=SUMAPRODUCTO(Costo,Asignación)

formato del modelo de asignación o de transporte. Por tanto, no tienen que dividirse las plantas 1 y 2 en dos asignados cada una, o agregarse tareas ficticias. En su lugar, las plantas 1 y 2 tienen un suministro de 2 cada una y después se introducen los signos ≤ en las celdas H19 y H20 al igual que las restricciones correspondientes en el cuadro de diálogo de Solver. Tampoco hay necesidad de incluir el método de la gran *M* para evitar la asignación del producto 3 a la planta 2 en la celda E20 ya que este cuadro de diálogo incluye la restricción de que E20 = 0. La celda objetivo CostoTotal (I24) muestra el costo total de $3 290 por día.

Ahora es de interés comparar esta solución con la que se obtuvo para la opción 1, que incluye la separación del producto 4 entre las plantas 2 y 3. Las asignaciones son diferentes en las dos soluciones, pero el costo total es casi el mismo: $3 260 en la opción 1 y $3 290 en la opción 2. Sin embargo, existen costos ocultos asociados con la separación de productos —que incluyen los costos de preparación adicional, de distribución y de administración—, los cuales no se incluyeron en la función objetivo para la opción 1. Igual que en cualquier aplicación de IO, el modelo matemático que se utiliza puede proporcionar sólo una representación aproximada del problema completo, de manera que la administración debe considerar factores que no se puedan incorporar al modelo antes de tomar su decisión final. En este caso, después de evaluar las desventajas de la separación de productos, la administración decidió adoptar la solución de la opción 2.

8.4 UN ALGORITMO ESPECIAL PARA EL PROBLEMA DE ASIGNACIÓN

En la sección 8.3 se puntualizó que el método símplex de transporte se puede usar para resolver problemas de asignación, pero que un algoritmo especial diseñado para tales problemas sería más eficiente. A continuación se describirá un algoritmo clásico de este tipo, llamado **algoritmo húngaro** (o *método húngaro*) porque fue desarrollado por matemáticos de Hungría. Se hará hincapié sólo en las ideas clave sin entrar en todos los detalles necesarios para llevar a cabo una implantación completa en computadora.

Papel de las tablas de costos equivalentes

El algoritmo opera en forma directa sobre la *tabla de costos* del problema. Con mayor precisión, convierte la tabla de costos original en una serie de tablas de costos *equivalentes* hasta llegar a una en la cual la solución óptima sea obvia. Esta tabla de costos equivalente consiste sólo en elementos *positivos* o iguales a *cero* donde todas las asignaciones se pueden hacer a las posiciones de los elementos con valor cero. Como el costo total no puede ser negativo, resulta claro que este conjunto de asignaciones con un costo total de cero es óptimo. La pregunta que resta es cómo convertir la tabla de costos original a esta forma.

La clave de esta conversión es la posibilidad de sumar o restar cualquier constante de cualquier elemento de un renglón o columna de la tabla de costos sin cambiar en realidad el problema, esto es, una solución óptima para la nueva tabla de costos también debe ser óptima para la antigua, y a la inversa.

Por tanto, el algoritmo comienza al restar el número más pequeño de cada renglón de todos los números del renglón. Este proceso de *reducción de renglón* creará una tabla de costos equivalente que tiene elementos iguales a cero en todos los renglones. Si esta tabla de costos tiene algunas columnas sin un elemento cero, el siguiente paso es realizar un proceso de *reducción de columna* al restar el número más pequeño de cada una de tales columnas del resto de los números de la columna.[12] La nueva tabla de costos equivalente tendrá un elemento cero en todos los renglones y columnas. Si los elementos cero proporcionan un conjunto completo de asignaciones, estas últimas constituyen una solución óptima y el algoritmo termina.

A manera de ilustración, considere la tabla de costos del problema de la Job Shop Co., que se proporcionó en la tabla 8.25. Para convertir esta tabla de costos en una tabla de costos equivalente,

[12] En realidad, los renglones y las columnas individuales se pueden reducir en cualquier orden, pero si se inicia con todos los renglones y después se continúa con todas las columnas se encuentra una forma sistemática de ejecutar el algoritmo.

suponga que se comienza el proceso de reducción de renglón al restar 11 de cada elemento del renglón 1, de donde se obtiene

	1	2	3	4
1	2	5	1	0
2	15	M	13	20
3	5	7	10	6
4(D)	0	0	0	0

Como cualquier solución factible debe tener exactamente una asignación en el renglón 1, el costo total de la nueva tabla siempre será exactamente 11 menos que el de la tabla antigua. Por tanto, la solución que minimiza el costo total de una tabla debe también minimizar el costo total de la otra.

Observe que mientras la tabla de costos original tenía sólo elementos estrictamente positivos en los primeros tres renglones, la nueva tabla tiene un elemento cero en el renglón 1. Como el objetivo es obtener elementos cero colocados en forma tan estratégica como para producir un conjunto completo de asignaciones, este proceso debería continuar en los otros renglones y columnas. Se deben evitar los elementos negativos, de manera que la constante que será restada debe ser el elemento mínimo del renglón o columna. Si se aplica este procedimiento a los renglones 2 y 3 se obtiene la siguiente tabla de costos equivalente:

	1	2	3	4
1	2	5	1	[0]
2	2	M	[0]	7
3	[0]	2	5	1
4(D)	0	[0]	0	0

Esta tabla de costos tiene todos los elementos cero que se requieren de un conjunto completo de asignaciones, como lo indican los cuatro recuadros, de manera que estas cuatro asignaciones constituyen una *solución óptima*, como se estableció en la sección 8.3 para este problema. El costo total de esta solución óptima se observa en la tabla 8.25, $Z = 29$, el cual es exactamente la suma de los números que se han restado de los renglones 1, 2 y 3.

Desafortunadamente, una solución óptima no siempre se obtiene con tanta facilidad, como se ejemplifica a continuación con la formulación del problema de asignación para la opción 2 de la Better Products Co., que se muestra en la tabla 8.29.

Como esta tabla de costos del problema tiene elementos cero en todos los renglones, suponga que se inicia el proceso de convertirla en tablas de costos equivalentes mediante la resta del elemento mínimo de cada columna de entrada. El resultado se muestra a continuación.

	1	2	3	4	5(D)
1a	80	0	30	120	0
1b	80	0	30	120	0
2a	60	60	M	80	0
2b	60	60	M	80	0
3	0	90	0	0	M

Ahora *todos* los renglones y columnas tienen al menos un elemento cero, pero esta vez *no* es posible un conjunto completo de asignaciones con elementos de este tipo. En realidad, el número máximo de asignaciones que puede hacerse en posiciones de elementos cero es de sólo 3. (Intente hacerlo.) Por tanto, se debe implantar una idea más compleja para terminar de resolver este problema, la cual no se requirió en el primer ejemplo.

Creación de elementos cero adicionales

Esta idea implica una nueva forma de crear posiciones *adicionales* con elementos cero sin crear ningún elemento negativo. En lugar de restar una constante a un *solo* renglón o columna, ahora se suma o se resta una constante a una *combinación* de renglones y columnas.

Este procedimiento se inicia dibujando un conjunto de líneas a través de algunos de los renglones y columnas de tal forma que se *cubran todos los ceros*. Es preferible que este procedimiento se haga con el número mínimo de líneas, como se muestra en la siguiente tabla de costos.

	1	2	3	4	5(*D*)
1*a*	80	0	30	120	0
1*b*	80	0	30	120	0
2*a*	60	60	*M*	80	0
2*b*	60	60	*M*	80	0
3	0	90	0	0	*M*

Observe que el elemento mínimo que no ha sido cruzado es 30, en las dos posiciones superiores de la columna 3. Por tanto, si se resta 30 de todos los elementos de la tabla completa, es decir, de cada renglón y columna, se creará un nuevo elemento cero en estas dos posiciones. Después, para restaurar los elementos cero previos y eliminar elementos negativos, se suma 30 a cada renglón y columna que esté cruzada por una línea, esto es, renglón 3 y columnas 2 y 5(*D*). De aquí se obtiene la siguiente tabla de costos equivalente.

	1	2	3	4	5(*D*)
1*a*	50	0	0	90	0
1*b*	50	0	0	90	0
2*a*	30	60	*M*	50	0
2*b*	30	60	*M*	50	0
3	0	120	0	0	*M*

Un atajo para obtener esta tabla de costos a partir de la anterior es restar 30 sólo de los elementos que no están cruzados por una línea y después sumar 30 a cada elemento que se encuentre en la intersección de dos líneas.

Observe que las columnas 1 y 4 de esta nueva tabla de costos tienen sólo un elemento cero y ambos están en el mismo renglón (renglón 3). En consecuencia, ahora es posible hacer cuatro asignaciones a posiciones de elementos cero, pero todavía no a cinco. (Intente hacerlo.) Por tanto, se repite el procedimiento anterior, donde ahora el número mínimo de líneas para cubrir todos los ceros es cuatro: la misma cantidad que el número máximo de asignaciones. Una forma de hacerlo es como se muestra a continuación.

	1	2	3	4	5(*D*)
1*a*	50	0	0	90	0
1*b*	50	0	0	90	0
2*a*	30	60	*M*	50	0
2*b*	30	60	*M*	50	0
3	0	120	0	0	*M*

De nuevo, el elemento mínimo no cubierto por una línea es 30, número que ahora aparece en la primera posición de los renglones 2*a* y 2*b*. Por tanto, se resta 30 de todos los elementos *no cubiertos*

y se suma 30 a todos los elementos *cubiertos dos veces* —sólo se pasan por alto los elementos de M—, lo cual proporciona la siguiente tabla de costos equivalente.

	1	2	3	4	5(*D*)
1*a*	50	[0]	0	90	30
1*b*	50	0	[0]	90	30
2*a*	[0]	30	*M*	20	0
2*b*	0	30	*M*	20	[0]
3	0	120	0	[0]	*M*

En realidad, con esta tabla existen algunas formas de hacer un conjunto completo de asignaciones a las posiciones de elementos cero (varias soluciones óptimas), incluso las que indican los cinco cuadros. El costo total resultante se observa en la tabla 8.29 como

$$Z = 810 + 840 + 800 + 0 + 840 = 3\,290.$$

A continuación se resume el algoritmo completo que se acaba de ejemplificar.

Resumen del algoritmo húngaro

1. Reste el número más pequeño de cada renglón a cada número del renglón. (Esto se llama *reducción de renglón*.) Introduzca los resultados en una nueva tabla.
2. Reste el número más pequeño de cada columna de la nueva tabla a cada número de la columna. (Esto se llama *reducción de columna*.) Introduzca los resultados en otra tabla.
3. Pruebe si se puede hacer una asignación óptima. Hágalo mediante la determinación del número mínimo de líneas necesario para cubrir (es decir, cruzar) todos los ceros. Puesto que este número mínimo de líneas es igual al número máximo de tareas que pueden realizarse en cero posiciones de elementos, si el número mínimo de líneas es igual al número de renglones, es posible tener un conjunto óptimo de tareas. (Si usted encuentra que no es posible un conjunto completo de tareas, esto significa que usted no redujo el número de líneas que cubren a todos los ceros hasta el número mínimo.) En ese caso vaya al paso 6. En caso contrario continúe con el paso 4.
4. Si el número de líneas es menor que el número de renglones, modifique la tabla de la siguiente forma:
   a) Reste el número no cubierto más pequeño de todos los números no cubiertos de la tabla.
   b) Sume el número no cubierto más pequeño a los números que se encuentran en las intersecciones de las líneas.
   c) Los números cruzados pero que no se encuentran en las intersecciones de las líneas permanecen sin cambio en la siguiente tabla.
5. Repita los pasos 3 y 4 hasta que sea posible tener un conjunto de asignaciones óptimo.
6. Haga las asignaciones una a una en las posiciones que tienen elementos cero. Comience con los renglones y columnas que tienen sólo un cero. Como cada renglón y cada columna necesita recibir exactamente una asignación, cruce tanto el renglón como la columna involucrados luego de hacer cada asignación. Después continúe con los renglones y columnas que aún no han sido cruzados para seleccionar la siguiente asignación, y de nuevo dé preferencia a algún renglón o columna que tenga un solo cero que no haya sido cruzado. Continúe hasta que todos los renglones y columnas tengan exactamente una asignación y por ende ya hayan sido cruzados.

En el IOR Tutorial se proporciona una rutina interactiva para aplicar este algoritmo de manera eficiente. También se incluye una rutina automática.

8.5 CONCLUSIONES

El modelo de programación lineal abarca una gran variedad de tipos específicos de problemas. El método símplex general es un algoritmo poderoso que puede resolver versiones muy grandes de

cualquiera de estos problemas. No obstante, algunos de ellos tienen formulaciones tan sencillas que se pueden resolver de manera mucho más eficiente mediante algoritmos *simplificados* que aprovechan su *estructura especial*. Estos algoritmos simplificados pueden reducir mucho el tiempo de computadora que requieren los problemas grandes y algunas veces permiten que sea factible resolver problemas muy grandes en una computadora. En particular, esto es cierto en el caso de los dos tipos de problemas de programación lineal estudiados en este capítulo, el de transporte y el de asignación. Ambos tipos tienen cierto número de aplicaciones en común, así que es importante reconocerlos cuando surjan y usar el mejor algoritmo disponible. Estos algoritmos especiales se incluyen en algunos paquetes de software para programación lineal.

En la sección 9.6 se examinará de nuevo la estructura especial de los problemas de transporte y asignación. Ahí se verá que estos problemas son casos especiales de una clase importante de problemas de programación lineal conocidos como *problemas de flujo de costo mínimo*, que se interpretan como minimizar el costo del flujo de bienes a través de una red. La versión simplificada del método símplex llamada *método símplex de redes* (descrito en la sección 9.7) se usa ampliamente para solucionar problemas de este tipo, e incluye varios casos especiales.

Un capítulo complementario (capítulo 23) del sitio de internet de este libro describe varios tipos especiales adicionales de problemas de programación lineal. Uno de ellos, llamado *problema de trasbordo*, es una generalización del problema de transporte que permite que los envíos de cualquier fuente a cualquier destino vayan primero a puntos de trasbordo. Como se trata de un caso especial del problema de flujo de costo mínimo, se describirá con más detalle en la sección 9.6.

Una gran parte de la investigación continúa dedicada al desarrollo de algoritmos simplificados para solucionar problemas de programación lineal de tipo especial, entre los que se cuentan algunos que no se presentaron aquí. Al mismo tiempo, existe un amplio interés en las aplicaciones de programación lineal para optimizar la operación de sistemas complicados a gran escala, incluso los sistemas sociales. Las formulaciones que resultan casi siempre tienen estructuras especiales que se puede aprovechar. La capacidad de reconocer y explotar las estructuras especiales se ha convertido en un factor muy importante en la aplicación exitosa de la programación lineal.

REFERENCIAS SELECCIONADAS

1. Dantzig, G. B. y M. N. Thapa, *Linear Programming 1: Introduction*, Springer, Nueva York, 1997, cap. 8.
2. Hall, R. W., *Handbook of Transportation Science*, 2a. ed., Kluwer Academic Publishers (ahora Springer), Boston, 2003.
3. Hillier, F. S. y M. S. Hillier, *Introduction to Management Science: A Modeling and Case Studies Approach with Spreadsheets*, 3a. ed., McGraw-Hill/Irwin, Burr Ridge, IL, 2003, cap. 6.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN EL SITIO DE INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos para el capítulo 8

Un ejemplo de demostración en OR Tutor:

El problema de transporte (*The Transportation Problem*)

Rutinas interactivas del IOR Tutorial:

Introducción o revisión de un problema de transporte (*Enter or Revise a Transportation Problem*)
Localización de una solución inicial básica factible, para el método interactivo (*Find Initial Basic Feasible Solution—for Interactive Method*)
Solución interactiva por el método símplex de transporte (*Solve Interactively by the Transportation Simplex Method*)
Solución interactiva de un problema de asignación (*Solve an Assignment Problem Interactively*)

Rutinas automáticas del IOR Tutorial:

Solución automática por el método símplex de transporte (*Solve Interactively by the Transportation Simplex Method*)
Solución automática de un problema de asignación (*Solve an Assignment Problem Automatically*)

Complemento de Excel:

Premium Solver for Education

Archivos para resolver los ejemplos "Ch.8—Transp. & Assignment":

Archivos de Excel
Archivo de LINGO/LINDO
Archivo de MPL/CPLEX

Glosario del capítulo 8

Suplemento de este capítulo:

Caso de estudio con muchos problemas de transporte

Vea el apéndice 1 para la documentación del software.

PROBLEMAS

Los símbolos a la izquierda de algunos problemas (o de sus incisos) significan:

D: El ejemplo de demostración puede ser útil.
I: Se sugiere al lector que use el procedimiento pertinente del IOR Tutorial, entre los que se indicó (la impresión registra su trabajo).
C: Use la computadora con cualquier opción de software disponible (o según indique el profesor) para resolver el problema.

Un asterisco en el número del problema indica que al final del libro se da al menos una respuesta parcial.

8.1-1. Lea el artículo de referencia que describe con detalle el estudio de OR que se resume en el Recuadro de aplicación de la sección 8.1. Allí se expone brevemente el modelo de transporte que se aplicó en ese estudio. Presente una lista con los diversos beneficios financieros y de otro tipo que generó dicho estudio.

8.1-2. La Compañía Childfair tiene tres plantas de producción de carros para bebés que deben distribuirse a cuatro centros de distribución. Las plantas 1, 2 y 3 producen 12, 17 y 11 cargamentos por mes, respectivamente. Cada centro de distribución necesita recibir 10 cargamentos por mes. En la siguiente tabla se da la distancia de cada planta a su respectivo centro de distribución:

		Distancia			
		Centro de distribución			
		1	2	3	4
	1	800 millas	1 300 millas	400 millas	700 millas
Planta	2	1 100 millas	1 400 millas	600 millas	1 000 millas
	3	600 millas	1 200 millas	800 millas	900 millas

El costo del flete de cada embarque es de $100 más 0.50 centavos por milla.

¿Cuánto se debería embarcar a cada centro de distribución para minimizar el costo total del envío?

a) Formule el problema como uno *de transporte* mediante la elaboración de una tabla de parámetros apropiada.
b) Trace la representación de red de este problema.
C *c*) Obtenga una solución óptima.

8.1-3.* Tom desearía comprar exactamente 3 litros de cerveza casera hoy y al menos 4 litros mañana. Dick quiere vender un máximo de 5 litros en total a un precio de $3.00 por litro hoy y de $2.70 por litro mañana. Harry está dispuesto a vender un máximo de 4 litros en total, a un precio de $2.90 por litro hoy y $2.80 por litro mañana.

Tom quiere saber cuánto debe comprar a cada uno para minimizar su costo y a la vez cumplir con los requisitos mínimos para satisfacer su sed.

a) Formule el modelo de *programación lineal* y elabore la tabla símplex inicial (vea los capítulos 3 y 4).
b) Formule este problema como un *problema de transporte* mediante la construcción de la tabla de parámetros apropiada.
C *c*) Obtenga una solución óptima para este problema.

8.1-4. La corporación Versatech producirá tres productos nuevos. En este momento, cinco de sus plantas tienen exceso de capacidad de producción. El costo unitario respectivo de fabricación del primer producto será de $41, $39, $42, $38 y $39 en las plantas 1, 2, 3, 4 y 5. El costo unitario de fabricación del segundo producto será de $55, $51, $56, $52 y $53 en las plantas respectivas 1, 2, 3, 4 y 5; y para el tercer producto será de $48, $45 y $50 en las plantas respectivas 1, 2 y 3, pero las plantas 4 y 5 no pueden fabricar este producto. Los pronósticos de ventas indican que la producción diaria debe ser de 700, 1 000 y 900 unidades de los productos 1, 2 y 3, respectivamente. Las

plantas 1, 2, 3, 4 y 5 tienen capacidades para producir 400, 600, 400, 600 y 1 000 unidades diarias, sin importar el producto o combinación de productos. Suponga que cualquier planta que tiene capacidad y posibilidad de fabricarlos podrá producir cualquier cantidad de productos y con cualquier combinación.

La administración desea asignar los nuevos productos a las plantas con el mínimo costo total de fabricación.

***a*)** Formule este problema como un *problema de transporte* mediante la construcción de la tabla de parámetros apropiada.

C ***b*)** Obtenga una solución óptima para este problema.

8.1-5. Reconsidere el problema de la P & T Co., que se presentó en la sección 8.1. Ahora se sabe que uno o más de los costos de envío por camión que se dio en la tabla 8.2 pueden cambiar antes de iniciar los embarques.

Use la herramienta Solver, de Excel, para generar el informe de sensibilidad. Utilice este informe para determinar el intervalo permisible para que cada costo unitario siga óptimo. ¿Qué indican estos intervalos permisibles a la administración de P & T?

8.1-6. La Onenote Co., que fabrica un solo producto, tiene tres plantas y cuatro clientes. Las plantas respectivas podrán producir 60, 80 y 40 unidades, durante el siguiente periodo. La empresa se ha comprometido a vender 40 unidades al cliente 1, 60 unidades al cliente 2 y por lo menos 20 unidades al cliente 3. Tanto el cliente 3 como el 4 desean comprar tantas unidades como sea posible de las restantes. La ganancia neta asociada con el envío de una unidad de la planta *i* al cliente *j* está dada en la tabla:

		Cliente			
		1	2	3	4
Planta	1	$800	$700	$500	$200
	2	$500	$200	$100	$300
	3	$600	$400	$300	$500

La administración desea saber cuántas unidades debe vender a los clientes 3 y 4, y cuántas unidades conviene enviar de cada planta a cada cliente, para maximizar la ganancia.

***a*)** Formule este problema como un problema de transporte donde la función objetivo sea maximizar mediante la construcción de la tabla de parámetros apropiada que proporcione la unidades de ganancia.

***b*)** Formule este problema de transporte con el objetivo de minimizar el costo total al convertir la tabla de parámetros del inciso *a*) en una con los costos unitarios en lugar de las ganancias unitarias.

***c*)** Presente la formulación de *a*) en una hoja de Excel.

C ***d*)** Use esta información y Excel Solver para obtener una solución óptima.

C ***e*)** Repita los incisos *c*) y *d*) para la formulación del inciso *b*). Compare las soluciones óptimas de las dos formulaciones.

8.1-7. La compañía Move-It tiene dos plantas que producen montacargas que se envían a tres centros de distribución. Los costos de producción unitarios son los mismos en las dos plantas y los costos de transporte por unidad de todas las combinaciones de planta y centro de distribución son los siguientes:

		Centro de distribución		
		1	2	3
Planta	A	$800	$700	$400
	B	$600	$800	$500

Se debe producir y enviar un total de 60 unidades por semana. Cada planta puede producir y enviar cualquier cantidad hasta un máximo de 50 unidades a la semana, de manera que hay una gran flexibilidad para dividir la producción total entre las dos plantas y reducir los costos de transporte.

El objetivo de la administración es determinar cuánto se debe producir en cada planta y, después, cuál debe ser el patrón de embarque de manera que se minimice el costo total de transporte.

***a*)** Formule este problema como un problema de transporte mediante la construcción de la tabla de parámetros apropiada.

***b*)** Presente el problema en una hoja de Excel.

C ***c*)** Use Solver, de Excel, para obtener una solución óptima.

8.1-8. Resuelva de nuevo el problema 8.1-7 si cualquier centro de distribución puede recibir cualquier cantidad entre 10 y 30 montacargas por semana para reducir más el costo total de envío, siempre que el envío total a los tres centros sea igual a 60 montacargas por semana.

8.1-9. La compañía MKJ debe producir una cantidad suficiente de dos artículos para cumplir con las ventas contratadas para los próximos tres meses. Los dos productos comparten las mismas instalaciones de producción y cada unidad de ambos requiere la misma capacidad de producción. Las capacidades de producción y almacenamiento disponibles cambian cada mes, por lo cual puede valer la pena producir más de alguno o ambos artículos en ciertos meses y almacenarlos hasta que se necesiten.

Para cada uno de los tres meses, la segunda columna de la siguiente tabla da el número máximo de unidades de los dos artículos combinados que se pueden producir en horas normales (HN) y en horas extra (HE). Para cada producto, las columnas subsecuentes dan 1) el número de unidades necesarias para la venta contratada, 2) el costo (en miles de dólares) por unidad en horas normales, 3) el costo (en miles de dólares) por unidad en horas extra y 4) el costo (en miles de dólares) de almacenar cada unidad adicional que se guarda para el siguiente mes. En cada caso, las cifras de los dos productos se separaron con una diagonal, con el valor del artículo 1 a la izquierda y el del artículo 2 a la derecha.

	Producción máxima combinada		Producto 1/Producto 2			
				Costo unitario de producción (miles de $)		Costo unitario de almacenamiento
Mes	HN	HE	Ventas	HN	HE	(miles de $)
1	10	3	5/3	15/16	18/20	1/2
2	8	2	3/5	17/15	20/18	2/1
3	10	3	4/4	19/17	22/22	

El gerente de producción quiere desarrollar un programa para elaborar el número de unidades de cada producto que debe fabricarse en horas normales y en horas extra —si se usa toda la capacidad de

producción normal— en cada uno de los tres meses. El objetivo es minimizar el costo total de producción y almacenamiento, mientras se cumple con las ventas contratadas para cada mes. No se tiene un inventario inicial y no se desea inventario final después de los 3 meses.

a) Formule este problema como un problema de transporte mediante la construcción de la tabla de parámetros apropiada.

C ***b***) Obtenga una solución óptima para este problema.

8.2-1. Considere el problema de transporte que tiene la siguiente tabla de costos y requerimientos:

		Destino			
		1	2	3	Recursos
	1	6	3	5	4
Origen	2	4	*M*	7	3
	3	3	4	3	2
Demanda		4	2	3	

a) Utilice el método de aproximación de Vogel —no use la rutina interactiva del IOR Tutorial— para elegir la primera variable básica de una solución inicial BF.

b) Utilice el método de aproximación de Russell a mano para elegir la primera variable básica de una solución inicial BF.

c) Utilice la regla de la esquina noroeste a mano para construir una solución inicial BF completa.

D,I **8.2-2.*** Considere el problema de transporte que tiene la siguiente tabla de parámetros:

		Destino					
		1	2	3	4	5	Recursos
	1	2	4	6	5	7	4
Origen	2	7	6	3	*M*	4	6
	3	8	7	5	2	5	6
	4	0	0	0	0	0	4
Demanda		4	4	2	5	5	

Utilice cada uno de los siguientes criterios para obtener una solución inicial BF.

Compare los valores de la función objetivo de estas soluciones.

a) Regla de la esquina noroeste.

b) Método de aproximación de Vogel.

c) Método de aproximación de Russell.

D,I **8.2-3.** Considere el problema de transporte cuya tabla de parámetros se muestra a continuación:

		Destino						
		1	2	3	4	5	6	Recursos
	1	13	10	22	29	18	0	5
	2	14	13	16	21	*M*	0	6
Origen	3	3	0	*M*	11	6	0	7
	4	18	9	19	23	11	0	4
	5	30	24	34	36	28	0	3
Demanda		3	5	4	5	6	2	

Utilice cada uno de los siguientes criterios para obtener una solución inicial BF. Compare los valores de la función objetivo de estas soluciones.

a) Regla de la esquina noroeste.

b) Método de aproximación de Vogel.

c) Método de aproximación de Russell.

8.2-4. Considere el problema de transporte cuya tabla de parámetros es:

		Destino				
		1	2	3	4	Recursos
	1	7	4	1	4	1
Origen	2	4	6	7	2	1
	3	8	5	4	6	1
	4	6	7	6	3	1
Demanda		1	1	1	1	

a) Observe que este problema tiene tres características especiales: 1) número de orígenes = número de destinos; 2) cada recurso = 1, y 3) cada demanda = 1. Los problemas de transporte con estas características son del tipo especial de problemas de asignación (descritos en la sección 8.3). Utilice la propiedad de soluciones enteras para explicar por qué este tipo de problemas de transporte se puede interpretar como la asignación de orígenes a destinos en una correspondencia uno a uno.

b) ¿Cuántas variables básicas hay en cada solución BF? ¿Cuántas son variables básicas degeneradas (= 0)?

D,I ***c***) Utilice la regla de la esquina noroeste para obtener la solución inicial BF.

I ***d***) Construya una solución inicial BF mediante la aplicación del procedimiento general del paso inicial del método símplex de transporte. Sin embargo, en lugar de usar uno de los tres criterios para el paso 1 que se presentó en la sección 8.2, utilice el criterio del costo mínimo que se presenta en seguida para seleccionar la siguiente variable básica. [Por medio de la rutina interactiva correspondiente del OR Courseware, elija la regla de la esquina noroeste (*Northwest Corner Rule*) debido a que esta elección permite utilizar cualquier criterio.]

> **Criterio de costo mínimo:** entre los renglones y columnas todavía en consideración, se elige la variable x_{ij} con el menor costo unitario c_{ij} como la siguiente variable básica (los empates se rompen de manera arbitraria).

D,I ***e***) A partir de la solución inicial BF del inciso *c*), aplique el método símplex de transporte en forma interactiva para obtener una solución óptima.

8.2-5. Considere el ejemplo prototípico del problema de transporte (de la P & T Co.) que se presentó al principio de la sección 8.1. Verifique que la solución óptima dada es de hecho óptima; para ello, aplique sólo la *prueba de optimalidad* del método símplex de transporte a esta solución.

8.2-6. Considere el problema de transporte que tiene la siguiente tabla de parámetros:

		Destino					
		1	2	3	4	5	Recursos
Origen	1	8	6	3	7	5	20
	2	5	M	8	4	7	30
	3	6	3	9	6	8	30
	4(D)	0	0	0	0	0	20
Demanda		25	25	20	10	20	

Después de varias iteraciones del método símplex de transporte, se obtienen las siguientes variables básicas: $x_{13} = 20, x_{21} = 25, x_{24} = 5, x_{32} = 25, x_{34} = 5, x_{42} = 0, x_{43} = 0, x_{45} = 20$. Continúe a mano con el método símplex de transporte *dos iteraciones más*. Después de ellas, establezca si la solución obtenida es óptima y por qué.

D,I **8.2-7.*** Considere el problema de transporte que tiene la siguiente tabla de parámetros:

		Destino				
		1	2	3	4	Recursos
Origen	1	3	7	6	4	5
	2	2	4	3	2	2
	3	4	3	8	5	3
Demanda		3	3	2	2	

Utilice cada uno de los siguientes criterios para obtener una solución inicial BF. En cada caso aplique en forma interactiva el método símplex de transporte, a partir de la solución inicial correspondiente, para obtener una solución óptima. Compare el número de iteraciones que se llevó a cabo.

a) Regla de la esquina noroeste.
b) Método de aproximación de Vogel.
c) Método de aproximación de Russell.

D,I **8.2-8.** La Cost-Less Corp., surte sus cuatro tiendas desde sus cuatro plantas. El costo de envío de cada planta a cada tienda se presenta en la siguiente tabla:

		Costo unitario de envío a cada tienda			
		1	2	3	4
Planta	1	\$700	\$800	\$500	\$200
	2	\$200	\$900	\$100	\$400
	3	\$400	\$500	\$300	\$100
	4	\$200	\$100	\$400	\$300

Las plantas respectivas 1, 2, 3 y 4 realizan 10, 20, 20 y 10 envíos al mes. Las tiendas 1, 2, 3 y 4 deben recibir 20, 10, 10 y 20 envíos respectivos por mes.

El gerente de distribución, Randy Smith, desea implementar el mejor plan para determinar cuántos envíos debe realizar de cada planta a cada tienda cada mes. El objetivo de Randy es minimizar el costo total de envío.

a) Formule este problema como uno de transporte mediante la construcción de la tabla de parámetros apropiada.
b) Use la regla de la esquina noroeste para desarrollar una solución BF inicial.
c) A partir de la solución básica inicial del inciso *b*), aplique el método símplex de transporte en forma interactiva para obtener una solución óptima.

8.2-9. La compañía Energetic debe planear el sistema de energía de un nuevo edificio.

Las necesidades de energía se refieren a las siguientes categorías: 1) electricidad, 2) calentadores de agua y 3) calefactores de ambiente. Los requerimientos diarios de energía (medidos en las mismas unidades) del edificio son:

Electricidad	30 unidades
Calentadores de agua	20 unidades
Calefactores de ambiente	50 unidades

Las tres fuentes posibles de energía son electricidad, gas natural y una unidad de celdas solares que se puede instalar en el techo. El tamaño de éste limita la cantidad de celdas solares a 30 unidades, pero no hay límite en la disponibilidad de electricidad y gas natural. Las necesidades de luz se pueden satisfacer sólo mediante la compra de energía eléctrica (a un costo de \$50 por unidad). Las otras dos necesidades se pueden satisfacer mediante cualquier fuente o combinación de fuentes. Los costos unitarios son

	Electricidad	Gas natural	Celdas solares
Calentadores de agua	\$150	\$110	\$70
Calefactores de ambiente	\$150	\$100	\$90

El objetivo es minimizar el costo total de satisfacer las necesidades de energía.

a) Formule este problema como uno de transporte mediante la construcción de la tabla de parámetros apropiada.
D,I ***b)*** Utilice la regla de la esquina noroeste para obtener una solución BF inicial para este problema.
D,I ***c)*** A partir de la solución inicial BF del inciso *b*), aplique en forma interactiva el método símplex de transporte para obtener una solución óptima.
D,I ***d)*** Utilice el método de aproximación de Vogel para obtener una solución BF inicial para este problema.
D,I ***e)*** A partir de la solución BF inicial que obtuvo en *d*), aplique en forma interactiva el método símplex de transporte para obtener una solución óptima.
I ***f)*** Utilice el método de aproximación de Russell para obtener una solución BF inicial para este problema.
D,I ***g)*** A partir de la solución que obtuvo en el inciso *f*) aplique en forma interactiva el método símplex de transporte para obtener una solución óptima. Compare el número de iteraciones que requirió este algoritmo con el número de iteraciones en los incisos *c*) y *e*).

D,I **8.2-10.*** Aplique en forma interactiva el método símplex de transporte para resolver el problema de programar la producción de la Northern Airplane Co., según se formuló en la tabla 8.9.

D,I **8.2-11.*** Reconsidere el problema 8.1-2.

a) Utilice la regla de la esquina noroeste para obtener una solución BF inicial.

b) A partir de la solución BF inicial del inciso *a*) aplique en forma interactiva el método símplex de transporte para obtener una solución óptima.

D,I **8.2-12.** Reconsidere el problema 8.1-3*b*. A partir de la regla de la esquina noroeste, aplique en forma interactiva el método símplex de transporte para obtener una solución óptima para este problema.

D,I **8.2-13.** Reconsidere el problema 8.1-4. A partir del método de aproximación de Vogel, aplique en forma interactiva el método símplex de transporte para obtener una solución óptima para este problema.

D,I **8.2-14.** Reconsidere el problema 8.1-6. Comience con el método de aproximación de Russell y aplique en forma interactiva el método símplex de transporte para obtener una solución óptima para este problema.

8.2-15. Reconsidere el problema de transporte que se presentó en el numeral 8.1-7*a*.

D,I ***a***) Utilice cada uno de los tres criterios que se presentaron en la sección 8.2 para obtener una solución BF inicial y anote el tiempo que emplea en cada una. Compare los tiempos y los valores de la función objetivo de estas soluciones.

C ***b***) Obtenga una solución óptima para este problema. Para cada solución BF inicial que obtuvo en el inciso *a*), calcule el porcentaje en el que el valor de la función objetivo excede el valor óptimo.

D,I ***c***) Para cada solución BF inicial que obtuvo en el inciso *a*), aplique el método símplex de transporte en forma interactiva para obtener (y verificar) una solución óptima. Anote el tiempo que emplea en cada caso. Compare los tiempos y el número de iteraciones necesarias para llegar a la solución óptima.

8.2-16. Siga las instrucciones del problema 8.2-15 para el modelo de transporte que se formuló en el problema 8.1-7*a*.

8.2-17. Considere el problema de transporte que tiene la siguiente tabla de parámetros:

		Destino		
		1	**2**	**Recursos**
Origen	1	8	5	4
	2	6	4	2
Demanda		3	3	

a) Elija un criterio de la sección 8.2 para obtener una solución BF inicial y resuelva el problema a mano mediante el método símplex de transporte. (Anote el tiempo empleado.)

b) Formule de nuevo este problema como un problema general de programación lineal y resuélvalo a mano mediante el *método símplex*. Mida el tiempo que le lleva resolverlo y compárelo con el tiempo que empleó en el inciso *a*).

8.2-18. Considere el problema de programar la producción de la Northern Airplane Co., de la sección 8.1 (vea la tabla 8.7). Formule este problema como un problema general de programación lineal en el que las variables de decisión son x_j = número de motores de turbina que se deben producir en el mes j (j = 1, 2, 3, 4). Construya la tabla símplex inicial de este modelo y después compare el tamaño de esta tabla (número de renglones y columnas) con el de la tabla símplex correspondiente a la formulación del problema de transporte (vea la tabla 8.9).

8.2-19. Considere la formulación de programación lineal general del problema de transporte (tabla 8.6). Verifique la afirmación de la sección 8.2 de que el conjunto de ($m + n$) ecuaciones de restricción —m restricciones de recursos y n de demanda— tiene una *ecuación redundante*; esto es, cualquier ecuación se puede expresar como una combinación lineal de las otras ($m + n - 1$) ecuaciones.

8.2-20. Cuando se manejan problemas de transporte donde las cantidades de recursos y de demanda son valores *enteros* explique por qué los pasos del método símplex de transporte garantizan que todas las variables básicas (asignaciones) de las soluciones BF que se obtienen presentan valores enteros. Comience por explicar por qué ocurre esto en el paso inicial cuando se usa el procedimiento general para construir una solución BF *inicial*, sin que importe el criterio para seleccionar la siguiente variable básica. Después, explique por qué, dada una solución BF *actual* entera, el paso 3 de una iteración debe obtener una solución BF que también lo sea. Por último, explique cómo se puede usar el paso inicial para construir *cualquier* solución BF inicial, de manera que el método símplex de transporte proporcione una prueba de la propiedad de soluciones enteras que se presentó en la sección 8.1.

8.2-21. Una contratista, Susan Meyer, tiene que acarrear grava a tres construcciones. Puede comprar hasta 18 toneladas a una cantera ubicada al norte de la ciudad y 14 toneladas a una del sur. Necesita 10, 5 y 10 toneladas en las respectivas construcciones 1, 2 y 3. El precio de compra por tonelada en cada cantera y los costos de acarreo son los siguientes:

	Costo por tonelada acarreada			
Cantera	**1**	**2**	**3**	**Precio por tonelada**
Norte	\$100	\$ 90	\$160	\$300
Sur	\$180	\$110	\$140	\$420

Susan desea determinar cuánto acarrear de cada cantera a cada construcción de manera que se minimice el costo total de compra y acarreo de la grava.

a) Formule el modelo de programación lineal. Use el método de la gran M para construir la tabla símplex inicial lista para aplicar el método símplex (pero no lo resuelva).

b) Luego formule este problema como uno de transporte mediante la construcción de la tabla de parámetros adecuada. Compare el tamaño de esta tabla —y de la tabla símplex de transporte correspondiente— que se utilizó por el método símplex de transporte, con el tamaño de la tabla símplex del inciso *a*) necesaria para aplicar el método símplex.

D ***c***) La contratista ha observado que puede abastecer por completo las construcciones 1 y 2 con material de la cantera norte y la construcción 3 con grava de la cantera sur. Utilice la prueba de optimalidad (pero no realice iteraciones) del método símplex de transporte para verificar si la solución BF correspondiente es óptima.

D,I ***d*)** A partir de la regla de la esquina noroeste, aplique el método símplex de transporte en forma interactiva para resolver el problema que se formuló en el inciso *b*).

***e*)** Como siempre, sea c_{ij} el costo unitario asociado con el origen *i* y el destino *j* dado en la tabla de parámetros construida en el inciso *b*). En el caso de la solución óptima que se obtuvo en el inciso *d*), suponga que el valor de c_{ij} de cada variable básica x_{ij} se fija en el valor dado en la tabla de parámetros, pero que el valor de c_{ij} de cada variable no básica x_{ij} tal vez se pueda alterar para buscar mejor precio, porque el gerente de la construcción quiere hacer crecer sus negocios. Aplique el análisis de sensibilidad para determinar el *intervalo permisible* para que cada valor x_{ij} anterior conserve su optimalidad y explique en qué puede ser útil esta información para la contratista.

C **8.2-22.** Considere la formulación del problema de transporte y la solución del problema del Metro Water District que se presentó en las secciones 8.1 y 8.2 (tablas 8.12 y 8.23).

Los números que se presentan en la tabla de parámetros son sólo estimaciones quizás inexactas, por lo que la administración desea hacer un análisis de "qué pasa si". Utilice el Excel Solver para generar un informe de sensibilidad para contestar las siguientes preguntas. (En cada caso, suponga que el cambio indicado es el único en el modelo.)

***a*)** ¿Conservaría su optimalidad la solución de la tabla 8.23 si el costo por acre-pie de agua enviada del río Calorie a San Go fuera de $200 en lugar de $230?

***b*)** ¿Seguiría óptima esta solución si el costo por acre-pie de agua enviada del río Sacron a Los Devils en realidad fuera de $160 en lugar de $130?

***c*)** ¿Debe seguir óptima esta solución si los costos considerados en *a*) y *b*) cambiaran al mismo tiempo a $215 y $145, respectivamente?

***d*)** Suponga que el suministro del río Sacron y la demanda de Hollyglass disminuyen en la misma cantidad. ¿Seguirán válidos los precios sombra para evaluar estos cambios si la disminución es de 0.5 millones de acre-pie?

8.2-23. Sin generar el informe de sensibilidad, adapte el procedimiento de las secciones 6.6 y 6.7 para realizar el análisis de sensibilidad especificado en los cuatro incisos del problema 8.2-22.

8.3-1. Considere un problema de asignación con la siguiente tabla de costos.

		Tarea			
		1	**2**	**3**	**4**
Asignado	A	8	6	5	7
	B	6	5	3	4
	C	7	8	4	6
	D	6	7	5	6

***a*)** Dibuje una representación de red de este problema.

***b*)** Formule este problema como uno de transporte mediante la construcción de la tabla de parámetros adecuada.

***c*)** Presente esta formulación un una hoja de Excel.

C ***d*)** Use Excel Solver para obtener una solución óptima.

8.3-2. Se deben utilizar cuatro barcos cargueros para transportar bienes de un puerto a otros cuatro puertos (numerados 1, 2, 3 y 4). Se puede usar cualquier barco para hacer cualquiera de los cuatro viajes. Sin embargo, dadas algunas diferencias entre las naves y las cargas, el costo total de carga, transporte y descarga de bienes de las distintas combinaciones de barcos y puertos varía de manera considerable. Estos costos se muestran en la siguiente tabla:

		Puerto			
		1	**2**	**3**	**4**
Barco	1	$500	$400	$600	$700
	2	$600	$600	$700	$500
	3	$700	$500	$700	$600
	4	$500	$400	$600	$600

El objetivo es asignar los barcos a los puertos en una correspondencia uno a uno de manera que se minimice el costo total de los cuatro envíos.

***a*)** Describa cómo se puede adecuar este problema al formato general de problemas de asignación.

C ***b*)** Obtenga una solución óptima.

***c*)** Reformule el problema como uno de transporte equivalente mediante la construcción de la tabla de parámetros apropiada.

D,I ***d*)** Use la regla de la esquina noroeste para obtener una solución BF inicial para el problema que se formuló en el inciso *c*).

D,I ***e*)** A partir de la solución BF inicial de *d*), utilice el método símplex de transporte para obtener un conjunto óptimo de asignaciones para el problema original.

D,I ***f*)** ¿Existen otras soluciones óptimas además de la que se obtuvo en el inciso *e*)? Si es así, utilice el método símplex de transporte para identificarlas.

8.3-3. Reconsidere el problema 8.1-4. Suponga que se revisaron los pronósticos de ventas y disminuyeron a 280, 400 y 350 unidades diarias de los respectivos productos 1, 2 y 3. De esta forma, cada planta tiene ahora capacidad para producir todo lo que se requiere de uno de ellos. La administración ha decidido que cada nuevo producto debe asignarse a una sola planta y que ninguna de éstas debe fabricar más de un producto (por ello, se asignará un producto a cada una de tres de las plantas y dos plantas quedarán sin asignación). El objetivo de hacer estas asignaciones es minimizar el *costo total* de producción de estas cantidades de los tres productos.

***a*)** Formule este problema como un problema de asignación mediante la construcción de la tabla de costos apropiada.

C ***b*)** Obtenga una solución óptima.

***c*)** Construya la tabla de parámetros adecuada para volver a formular el problema como uno de transporte.

D,I ***d*)** Comience con el método de aproximación de Vogel y aplique el método símplex de transporte en forma interactiva para resolver el problema que formuló en *c*).

8.3-4.* El entrenador de un equipo de natación debe asignar competidores para la prueba de 200 metros de relevo combinado que irá a las Olimpiadas Juveniles. Como muchos de sus mejores nadadores son rápidos en más de un estilo, no es fácil decidir cuál de ellos asignar a cada uno de los cuatro estilos. Los cinco mejores nadadores y sus mejores tiempos (en segundos) en cada estilo son los siguientes:

Tipo de nado	Carl	Chris	David	Tony	Ken
Dorso	37.7	32.9	33.8	37.0	35.4
Pecho	43.4	33.1	42.2	34.7	41.8
Mariposa	33.3	28.5	38.9	30.4	33.6
Libre	29.2	26.4	29.6	28.5	31.1

El entrenador quiere determinar cómo asignar cuatro nadadores a los cuatro estilos de nado para minimizar la suma de los mejores tiempos correspondientes

a) Formule este problema como uno de asignación.
C ***b***) Obtenga una solución óptima.

8.3-5. Considere el problema de asignación que se formuló para la opción 2 del problema de la Better Products Co., en la tabla 8.29.

a) Reformule este problema como un problema de transporte equivalente con tres orígenes y cinco destinos mediante la construcción de la tabla de parámetros apropiada.
b) Convierta la solución óptima que se obtuvo en la sección 8.3 para este problema de asignación en una solución BF completa —que incluya variables básicas degeneradas— para el problema de transporte formulado en *a*). En particular, aplique el "procedimiento general para construir una solución BF inicial" de la sección 8.2. Para cada iteración del procedimiento, en lugar de usar cualquiera de los tres criterios alternativos que se presentaron en el paso 1, seleccione la siguiente variable básica de modo que corresponda a la siguiente asignación de una planta a un producto en la solución óptima. Cuando quede sólo una columna o un renglón, utilice el paso 4 para elegir las variables básicas que faltan.
c) Verifique que la solución óptima de la sección 8.3 para este problema sea óptima, sólo con la prueba de optimalidad del método símplex de transporte sobre la solución BF completa del inciso *b*).
d) Luego reformule este problema de asignación como uno de transporte equivalente con cinco orígenes y cinco destinos mediante la construcción de la tabla de parámetros apropiada. Compare este problema con el del inciso *a*).
e) Repita el inciso *b*) para el problema formulado en *d*). Compare la solución BF obtenida con la del inciso *b*).

D,I **8.3-6.** A partir del método de aproximación de Vogel, aplique la rutina interactiva del método símplex de transporte para resolver el problema de asignación de la Job Shop Co., que se formuló en la tabla 8.26*b*). (Como se vio en la sección 8.3, la solución óptima que resulta tiene como variables básicas $x_{14} = 1$, $x_{23} = 1$, $x_{31} = 1$, $x_{42} = 1$, y el resto de las $x_{ij} = 0$.)

8.3-7. Reconsidere el problema 8.1-7. Suponga que los respectivos centros de distribución 1, 2 y 3 deben recibir exactamente 10, 20 y 30 unidades por semana. Por conveniencia administrativa, la gerencia ha decidido que cada centro de distribución sea abastecido por completo por una sola planta, de forma que una planta abastecerá un centro y la otra, los otros dos. La elección de estas asignaciones de plantas a centros de distribución debe hacerse basada sólo en la minimización del costo total de transporte.

a) Formule este problema como un problema de asignación mediante la construcción de la tabla de costos adecuada e identifique los asignados y las tareas correspondientes.
C ***b***) Obtenga una solución óptima.
c) Reformule este problema de asignación como un problema de transporte equivalente (con cuatro orígenes) mediante la construcción de la tabla de parámetros apropiada.
C ***d***) Resuelva el problema que se formuló en el inciso *c*).
e) Repita el inciso *c*) con sólo dos orígenes.
C ***f***) Resuelva el problema que se formuló en el inciso *e*).

8.3-8. Considere el problema de asignación que tiene la siguiente tabla de costos.

		Tarea		
		1	2	3
	A	5	7	4
Persona	B	3	6	5
	C	2	3	4

La solución óptima es A-3, B-1, C-2, con $Z = 10$.

C ***a***) Utilice la computadora para verificar esta solución óptima.
b) Reformule este problema como uno de transporte equivalente mediante la construcción de la tabla de parámetros adecuada.
C ***c***) Obtenga una solución óptima para el problema de transporte que se formuló en el inciso *b*).
d) ¿Por qué la solución BF óptima que se obtuvo en el inciso *c*) incluye algunas variables básicas (degeneradas) que no son parte de la solución óptima del problema de asignación?
e) Ahora considere las variables *no básicas* de la solución BF óptima que se obtuvo en el inciso *c*). Para cada variable no básica x_{ij} y el costo correspondiente c_{ij}, adapte el procedimiento de análisis de sensibilidad para programación lineal general —vea el caso 2*a* de la sección 6.7— para determinar el *intervalo de valores permisibles de* c_{ij}.

8.3-9. Considere el modelo de programación lineal del problema de asignación que se presentó en la sección 8.3. Construya la tabla de coeficientes de las restricciones de este modelo. Compare esta tabla con la del problema general de transporte (tabla 8.6). ¿En qué consiste la estructura especial del problema de asignación comparado con el problema general de transporte?

I **8.4-1.** Reconsidere el problema de asignación presentado en el problema 8.3-2. Aplique en forma manual el algoritmo húngaro para resolver este problema. (Puede usar la rutina interactiva correspondiente del IOR Tutorial.)

I **8.4-2.** Reconsidere el problema 8.3-4. Vea su formulación como un problema de asignación en las respuestas que se proporcionan al final del libro. Aplique en forma manual el algoritmo húngaro para resolver este problema. (Puede usar la rutina interactiva correspondiente del IOR Tutorial.)

I **8.4-3.** Reconsidere el problema de asignación que se formuló para la opción 2 del problema de la Better Products Co., en la tabla 8.29. Suponga que el costo de que la planta 1 fabrique el producto 1 se reduce de 820 a 720. Resuelva este problema por medio de la aplicación manual del algoritmo húngaro. (Puede usar la rutina interactiva correspondiente del IOR Tutorial.)

I **8.4-4.** Aplique en forma manual el algoritmo húngaro —quizá mediante el uso de la rutina interactiva correspondiente del IOR Tutorial— para resolver el problema de asignación que tiene la siguiente tabla de costos:

		Tarea		
		1	2	3
Persona	1	*M*	8	7
	2	7	6	4
	3(*D*)	0	0	0

I **8.4-5.** Aplique en forma manual el algoritmo húngaro —quizá mediante el uso de la rutina interactiva correspondiente del IOR Tutorial— para resolver el problema de asignación que tiene la siguiente tabla de costos:

		Tarea			
		1	2	3	4
Asignado	A	4	1	0	1
	B	1	3	4	0
	C	3	2	1	3
	D	2	2	3	0

I **8.4-6.** Aplique en forma manual el algoritmo húngaro —quizá mediante el uso de la rutina interactiva correspondiente del IOR Tutorial— para resolver el problema de asignación que tiene la siguiente tabla de costos:

		Tarea			
		1	2	3	4
Asignado	A	4	6	5	5
	B	7	4	5	6
	C	4	7	6	4
	D	5	3	4	7

CASOS

CASO 8.1 Envío de madera al mercado

Alabama Atlantic es una maderera que tiene tres fuentes de suministro de madera y cinco mercados que surtir. La disponibilidad anual de madera en los orígenes respectivos l, 2 y 3 es l5, 20 y l5 millones de pies lineales. La cantidad que se puede vender cada año en los mercados 1, 2, 3, 4 y 5 es 11, 12, 9, 10 y 8 millones de pies lineales, respectivamente.

En el pasado, la compañía ha enviado madera por tren. Sin embargo, debido a que los costos de flete han aumentado, se investiga la alternativa de usar barcos para algunas entregas. Esta alternativa requiere que la compañía invierta en algunos barcos. Excepto por estos costos de inversión, los costos de envío en miles de dólares por millón de pies lineales por tren y por agua (cuando sea factible) son los siguientes:

	Costo unitario por tren (miles de $) Mercado					Costo unitario por barco (miles de $) Mercado				
Origen	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
1	61	72	45	55	66	31	38	24	—	35
2	69	78	60	49	56	36	43	28	24	31
3	59	66	63	61	47	—	33	36	32	26

La inversión de capital en los barcos (en miles de dólares) requerida por cada millón de pies lineales transportados al año junto con cada ruta es:

	Inversión en barcos (miles de $) Mercado				
Origen	**1**	**2**	**3**	**4**	**5**
1	275	303	238	—	285
2	293	318	270	250	265
3	—	283	275	268	240

Al considerar la vida útil de los barcos y el valor del dinero en el tiempo, el costo anual equivalente de estas inversiones es un décimo de la cantidad que se muestra en la tabla. El objetivo es determinar el plan de envíos global que minimice el costo anual uniforme equivalente (incluidos los costos de envío).

Usted es el jefe del equipo de IO al que se asignó la tarea de determinar este plan de envío de cada una de las tres opciones siguientes:

Opción 1: Continuar con envíos sólo por ferrocarril.
Opción 2: Cambiar a envíos sólo por agua (excepto donde sólo el tren es factible).
Opción 3: Enviar por tren o por barco, de acuerdo con el menor costo de la ruta específica.

Presente sus resultados de cada opción. Haga una comparación.

Por último, considere el hecho de que estos resultados se basan en los costos *actuales* de envío e inversión, de manera que la decisión acerca de cuál opción elegir debe tomar en cuenta la proyección de la administración del cambio probable de estos costos en el futuro. Para cada opción, describa un escenario de cambios en los costos futuros que justificaría adoptar esa nueva opción.

(*Nota:* Para su conveniencia, los archivos de datos de este caso se proporcionan en el sitio de internet de este libro.)

RESUMEN DE CASOS ADICIONALES EN EL SITIO DE INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

CASO 8.2 Continuación del caso de estudio Texago

En el complemento de este capítulo incluido en el sitio de internet de este libro se presenta el caso de estudio sobre la forma en que la Texago Corp., resolvió muchos problemas de transporte para ayudar a su decisión relativa a dónde localizar su nueva refinería de petróleo. Ahora, la administración debe decidir si la capacidad de la refinería debería ser algo más grande que lo planeado en un inicio. Para ello debe formular y resolver algunos problemas de transporte adicionales. En consecuencia, una parte importante del análisis implicará la combinación de dos problemas de transporte en un solo modelo de programación lineal que en forma simultánea considere el envío de petróleo crudo de los pozos a las refinerías y el envío de producto final de las refinerías a los centros de distribución. También se requiere escribir una síntesis dirigida a la administración que resuma sus resultados y recomendaciones.

CASO 8.3 Elección de proyectos

Este caso se enfoca en una serie de aplicaciones del problema de asignación de una compañía que fabrica productos farmacéuticos. Se ha decidido considerar cinco proyectos de investigación y desarrollo para intentar el desarrollo de nuevos productos que servirán para tratar cinco tipos específicos de enfermedades. Para dirigir estos proyectos se encuentran disponibles cinco científicos reconocidos. El problema actual es decidir cómo asignar estos científicos a cada uno de los proyectos. En este caso se debe considerar una variedad de escenarios probables.

CAPÍTULO 9

Modelos de optimización de redes

Los problemas de redes surgen en una gran variedad de situaciones. Las redes de transporte, eléctricas y de comunicaciones predominan en la vida diaria. La representación de redes se utiliza de manera amplia en áreas tan diversas como producción, distribución, planeación de proyectos, localización de instalaciones, administración de recursos y planeación financiera, por mencionar sólo algunos ejemplos. En realidad, una representación de redes proporciona un poderoso apoyo visual y conceptual para mostrar las relaciones entre las componentes de los sistemas, de tal modo que se usa casi en todos los ámbitos científicos, sociales y económicos.

Uno de los mayores desarrollos recientes en investigación de operaciones (IO) ha sido el rápido avance tanto en la metodología como en la aplicación de los modelos de optimización de redes. La aparición de algunos algoritmos ha tenido un efecto importante, al igual que las ideas de ciencias de la computación acerca de estructuras de datos y la manipulación eficiente de éstos. En la actualidad se dispone de algoritmos y paquetes de computadora que *se usan* en forma rutinaria para resolver problemas muy grandes que no se habrían podido manejar hace dos o tres décadas.

Muchos modelos de optimización de redes son en realidad tipos especiales de problemas de *programación lineal*. Por ejemplo, tanto el problema de transporte como el de asignación, que se presentaron en el capítulo anterior, pertenecen a esta categoría debido a su representación mediante una red que se mostró en las figuras 8.3 y 8.5.

Uno de los ejemplos de programación lineal que se presentó en la sección 3.4 también es un problema de optimización de redes. Éste es el ejemplo de la Distribution Unlimited Co., que desea saber cómo repartir sus bienes en la red de distribución que se muestra en la figura 3.13. Este tipo especial de problema de programación lineal, llamado de *flujo de costo mínimo* se presenta en la sección 9.6. Se volverá a analizar este ejemplo en particular en esa sección y después se resolverá con la metodología de redes en la sección siguiente.

En este capítulo sólo serán planteadas las bases de la metodología de redes actual. Sin embargo, se presentará una introducción a cinco tipos importantes de problemas de redes y algunas ideas básicas sobre cómo resolverlos (sin profundizar en los aspectos de estructuras de bases de datos, tan vitales para la aplicación exitosa en los problemas a gran escala). Los tres primeros tipos de problemas —*el de la ruta más corta, el del árbol de mínima expansión y el del flujo máximo*— tienen una estructura específica que surge con frecuencia en la práctica.

El cuarto tipo —*problema del flujo de costo mínimo*— proporciona un enfoque unificado de muchas otras aplicaciones debido a su estructura mucho más general. Esta estructura es tan general que incluye como casos especiales el problema de la ruta más corta y el de flujo máximo, al igual que los problemas de transporte y de asignación del capítulo 8. En razón de que el problema del flujo de costo mínimo es un tipo especial de problema de programación lineal, se puede resolver en forma eficiente mediante una versión simplificada del método símplex llamada *método símplex de redes*. (No se presentarán problemas de redes aún más generales cuya solución es más complicada.)

El quinto tipo de problemas de redes que se considera aquí implica la determinación del modo más económico de realizar un proyecto de forma que éste pueda terminarse en su fecha límite. Se utiliza una técnica llamada *método CPM de trueques entre tiempo y costo* para formular un

modelo de red del proyecto y los trueques entre tiempo y costo para sus actividades. Después se utiliza el análisis de costo marginal o la programación lineal para resolver el plan de proyecto óptimo.

En la primera sección se introduce un ejemplo prototípico que se usará más adelante para ilustrar el enfoque de los primeros tres tipos de problemas. En la sección 9.2 se presenta la terminología básica de redes. Las siguientes cuatro secciones están dedicadas a cuatro tipos de problemas. En la sección 9.7 se estudia el método símplex de redes. Por último, en la sección 9.8 se presenta el método CPM de trueques entre tiempo y costo.

9.1 EJEMPLO PROTOTÍPICO

En fecha reciente se reservó el área de SEERVADA PARK para paseos y campamentos. No se permite la entrada de automóviles, pero existe un sistema de caminos angostos y sinuosos para tranvías y para "jeeps" conducidos por los guardabosques. En la figura 9.1 se muestra este sistema de caminos —sin las curvas—, en donde *O* es la entrada al parque; las otras letras representan la localización de las casetas de los guardabosques y otras instalaciones de servicio. Los números son las distancias en millas de estos caminos accidentados.

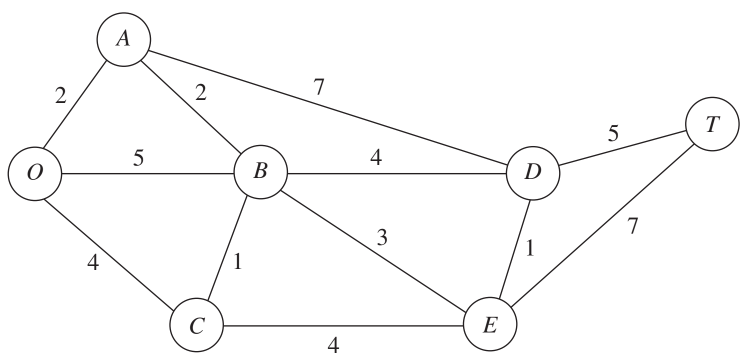
El parque contiene un mirador a un hermoso paisaje en la estación *T*. Unas cuantas camionetas transportan a los visitantes desde la entrada a la estación *T* y viceversa.

En este momento la administración del parque se enfrenta a tres problemas. Uno consiste en determinar qué ruta, desde la entrada del parque a la estación *T*, es la que representa la *distancia total más corta* para la operación de los tranvías. (Éste es un ejemplo del problema de la ruta más corta que se estudiará en la sección 9.3.)

El segundo problema consiste en que deben instalarse líneas telefónicas subterráneas para establecer comunicación entre todas las estaciones, incluso la entrada. Debido a que la instalación es cara y perturba la ecología, se deben instalar líneas que sigan sólo los caminos necesarios para obtener comunicación entre cualquier par de estaciones. La pregunta es por dónde deben tenderse las líneas para lograr este objetivo con el *mínimo* número total de millas de cable instalado. (Éste es un ejemplo del problema del árbol de mínima expansión que se analizará en la sección 9.4.)

El tercer problema se refiere a que, durante la temporada pico, hay más personas que quieren tomar un tranvía a la estación *T* que aquellas a las que se les puede dar servicio. Para evitar la perturbación indebida de la ecología y de la vida silvestre de la región, se ha impuesto un racionamiento estricto al número de viajes al día que pueden hacer los tranvías en cada camino. (Estos límites difieren según los caminos, como se verá con detalle en la sección 9.5.) De esta forma, durante la temporada pico, se pueden seguir varias rutas, sin tomar en cuenta la distancia, para aumentar el número de viajes de tranvía diarios. La pregunta es cómo planear las rutas de los distintos viajes, de manera que se *maximice* el número total de viajes que se pueden hacer al día, sin violar los límites impuestos sobre cada camino. (Éste es un ejemplo del problema de flujo máximo que se presentará en la sección 9.5.)

FIGURA 9.1
Sistema de caminos del Seervada Park.



9.2 TERMINOLOGÍA DE REDES

Se ha desarrollado una terminología relativamente extensa para describir los tipos de redes y sus componentes. Aunque se ha evitado en lo posible el uso del vocabulario específico, es necesario introducir un número considerable de términos que se usarán en este capítulo. Se sugiere al lector que lea la sección completa una vez para entender las definiciones y planee después regresar para repasar los términos a medida que éstos se utilicen en las secciones siguientes. Como ayuda, se resalta el nombre de cada término en **negritas** en el punto en que se define.

Una red consiste en un conjunto de *puntos* y un conjunto de *líneas* que unen ciertos pares de puntos. Los puntos se llaman **nodos** (o vértices); por ejemplo, la red de la figura 9.1 tiene siete nodos que son representados por siete círculos. Las líneas se llaman **arcos** (o ligaduras, aristas o ramas); por ejemplo, la red de la figura 9.1 tiene 12 arcos que corresponden a los 12 caminos del sistema del parque. Los arcos se etiquetan al dar el nombre de los nodos en sus puntos terminales; por ejemplo, en la figura 9.1, *AB* es el arco entre los nodos *A* y *B*.

Los arcos de una red pueden tener un flujo de algún tipo que pase por ellos; por ejemplo, el flujo de tranvías sobre los caminos de Seervada Park en la sección 9.1. La tabla 9.1 proporciona varios ejemplos de flujo en redes. Si el flujo a través de un arco se permite sólo en una dirección —como en una calle de un sentido—, se dice que el arco es un **arco dirigido**. La dirección se indica al agregar una cabeza de flecha al final de la línea que representa el arco. Cuando se etiqueta un arco dirigido con el nombre de los nodos que unen, siempre se pone primero el nodo *de donde viene* y después el nodo *hacia donde* va, esto es, un arco dirigido *del* nodo *A* al nodo *B* debe etiquetarse como *AB* y no como *BA*. Otra manera de etiquetarlo es $A \rightarrow B$.

Si el flujo a través de un arco se permite en ambas direcciones —como una tubería que se puede usar para bombear fluido en ambas direcciones—, se dice que el arco es un **arco no dirigido**. Para ayudar a distinguir entre los dos tipos de arcos, con frecuencia se hará referencia a los arcos no dirigidos con el sugestivo nombre de **ligadura**.

Aunque se permita que el flujo a través de un arco no dirigido ocurra en cualquier dirección, se supone que ese flujo será sólo en una dirección, en la seleccionada, y no se tendrán flujos simultáneos en direcciones opuestas. (Este último caso requiere usar *un par de arcos dirigidos* en direcciones opuestas.) En el proceso de toma de decisiones sobre el flujo de un arco no dirigido, se permite hacer una secuencia de asignaciones de flujos en direcciones opuestas, pero en el entendido de que el flujo real será el *flujo neto*, esto es, la diferencia de los flujos asignados en las dos direcciones. Por ejemplo, si se asigna un flujo de 10 en una dirección y después un flujo de 4 en la dirección opuesta, el efecto real es la *cancelación* de 4 unidades de la asignación original, lo que reduce el flujo en la dirección original de 10 a 6. Aun en el caso de un arco dirigido, en ocasiones se usa la misma técnica como una manera conveniente de reducir un flujo asignado con anterioridad. En particular, se puede hacer una asignación ficticia de flujo en la dirección “equivocada” a través de un arco dirigido para registrar una reducción en esa cantidad del flujo que va en la dirección “correcta”.

Una red que tiene sólo arcos dirigidos se llama **red dirigida**. De igual manera, si todos sus arcos son no dirigidos, se dice que se trata de una **red no dirigida**. Una red con una mezcla de arcos dirigidos y no dirigidos —o incluso una con todos sus arcos no dirigidos— se puede convertir en una red dirigida, si se desea, mediante la sustitución de cada arco no dirigido por un par de arcos dirigidos en direcciones opuestas. (Después se puede optar por interpretar los flujos a través de cada par de arcos dirigidos como flujos simultáneos en direcciones opuestas o de proporcionar un flujo neto en una dirección, según convenga al caso.)

TABLA 9.1 Componentes de redes representativas

Nodos	Arcos	Flujo
Cruceros	Caminos	Vehículos
Aeropuertos	Líneas aéreas	Aviones
Puntos de conmutación	Cables, canales	Mensajes
Estaciones de bombeo	Tuberías	Fluidos
Centros de trabajo	Rutas de manejo de materiales	Trabajos

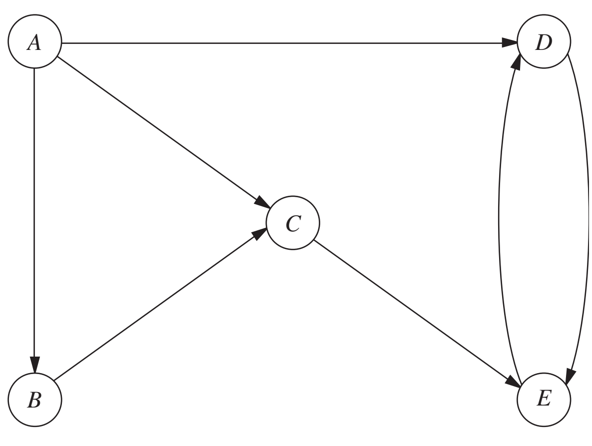
Cuando dos nodos no están unidos por un arco es válido preguntar si están conectados por una serie de arcos. Una **trayectoria** entre dos nodos es *una sucesión de arcos distintos* que conectan estos nodos. Por ejemplo, una de las trayectorias que conectan a los nodos O y T en la figura 9.1 es la sucesión de arcos OB–BD–DT ($O \to B \to D \to T$), y viceversa. Cuando algunos o todos los arcos de una red son arcos dirigidos, se hace la distinción entre trayectorias dirigidas y trayectorias no dirigidas. Una **trayectoria dirigida** del nodo i al nodo j es una sucesión de arcos cuya dirección (si la tienen) es *hacia* el nodo j, de manera que el flujo del nodo i al nodo j a través de esta trayectoria es factible. Una **trayectoria no dirigida** del nodo i al nodo j es una sucesión de arcos cuya dirección (si la tiene) puede ser *hacia o desde* el nodo j. (Observe que una trayectoria dirigida también satisface la definición de trayectoria no dirigida, pero el inverso no se cumple.) Con frecuencia, una trayectoria no dirigida tendrá algunos arcos dirigidos hacia el nodo j y otros desde él, es decir, hacia el nodo i. En las secciones 9.5 y 9.7 se verá que, tal vez de manera sorprendente, las trayectorias *no dirigidas* juegan un papel muy importante en el análisis de las redes *dirigidas*.

Para ilustrar estas definiciones, la figura 9.2 muestra una red dirigida común. (Sus nodos y arcos son los mismos que los de la figura 3.13, donde los nodos A y B representan dos fábricas y los nodos D y E representan dos almacenes, el nodo C es un centro de distribución y los arcos representan las rutas de embarque.) La sucesión de arcos AB–BC–CE es una trayectoria dirigida ($A \to B \to C \to E$) del nodo A al nodo E, puesto que el flujo hacia el nodo E en toda esta trayectoria es factible. Por otro lado, BC–AC–AD ($B \to C \to A \to D$) *no* es una trayectoria dirigida del nodo B al nodo D, porque la dirección del arco AC es desde el nodo D (sobre esta trayectoria). No obstante, $B \to C \to A \to D$ *es* una trayectoria no dirigida del nodo B al nodo D, debido a que la secuencia de arcos BC–AC–AD *conecta* a estos dos nodos (aun cuando la dirección del arco AC evita el flujo a través de esta trayectoria).

Como ejemplo de la relevancia de las trayectorias no dirigidas, suponga que se habían asignado dos unidades de flujo del nodo A al nodo C al arco AC. En razón de esta asignación previa, ahora es factible asignar un flujo más pequeño, por ejemplo una unidad, a la trayectoria no dirigida $B \to C \to A \to$ D, aunque la dirección de AC evite un flujo positivo a través de $C \to A$. La razón es que esta asignación de flujo en la dirección "equivocada" para el arco AC en realidad sólo *reduce* el flujo en la dirección "correcta" en una unidad. Las secciones 9.5 y 9.7 hacen un uso amplio de esta técnica de asignación de flujos a través de una trayectoria no dirigida que incluye arcos cuya dirección es opuesta al flujo, y en la que el efecto real sobre estos arcos es una reducción de los flujos positivos asignados antes en la dirección "correcta".

Un **ciclo** es una trayectoria que comienza y termina en el mismo nodo. En una red *dirigida*, un ciclo puede ser dirigido o no dirigido, según la trayectoria en cuestión sea dirigida o no dirigida. (Como una trayectoria dirigida también es no dirigida, un ciclo dirigido es un ciclo no dirigido, pero en general el inverso no es cierto.) Por ejemplo, en la figura 9.2, DE–ED es un ciclo dirigido. Por el contrario, AB–BC–AC no es un ciclo dirigido puesto que la dirección del arco AC es opuesta a la de los arcos AB y BC. Por otro lado, AB–BC–AC no es un ciclo dirigido porque $A \to B \to C \to A$

■ **FIGURA 9.2**
La red de distribución de Distribution Unlimited Co., que se presentó en la figura 3.13, ilustra una red dirigida.



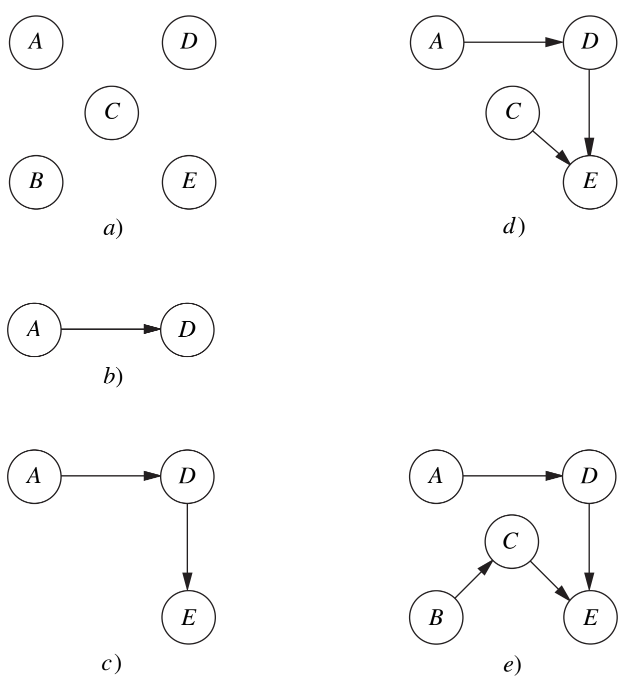
es una trayectoria no dirigida. En la red no dirigida que se muestra en la figura 9.1 existen muchos ciclos; por ejemplo, *OA–AB–BC–CO*. De cualquier forma, observe que la definición de *trayectoria* —una sucesión de arcos *distintos*— elimina la posibilidad de retroceder para formar un ciclo. Por ejemplo, *OB–BO* en la figura 9.1 no califica como ciclo, porque *OB* y *BO* son dos etiquetas del *mismo* arco (ligadura). Por otra parte, en la figura 9.2, *DE–ED* es un ciclo (dirigido) porque *DE* y *ED* son arcos distintos.

Se dice que dos nodos están **conectados** si la red contiene al menos una trayectoria *no dirigida* entre ellos. (Observe que no es necesario que la trayectoria sea dirigida aun cuando la red sea dirigida.) Una **red conexa** es una red en la que cada par de nodos está conectado. Entonces, las redes de las figuras 9.1 y 9.2 son ambas conexas. La última red no sería conexa si se eliminaran los arcos *AD* y *CE*.

Considere una red conexa con n nodos —por ejemplo, los $n = 5$ nodos de la figura 9.2— en la que han sido eliminados todos los arcos. Se puede “hacer crecer” un “árbol” si se agrega un arco —o “rama”— a la vez a partir de la red original de cierta manera. El primer arco puede ir en cualquier lugar de modo que conecte algún par de nodos. De ahí en adelante, cada arco nuevo debe agregarse entre un nodo que ya haya sido conectado a otros nodos y a un nuevo nodo no conectado. Si se agregan arcos de esta manera, se evita que se forme un ciclo y además se asegura que el número de nodos conexos sea uno más que el número de arcos. Cada nuevo arco crea un **árbol** más grande, que es una *red conexa* —para algún subconjunto de n nodos— que no contiene *ciclos no dirigidos*. Una vez agregado el $(n-1)$-ésimo arco, el proceso se detiene porque el árbol resultante se expande (conecta) hacia todos los n nodos. Este árbol, que se llama **árbol de expansión**, es una *red conexa* de los n nodos que contienen *ciclos no dirigidos*. Todo árbol de expansión tiene exactamente $n-1$ arcos, puesto que éste es el número *mínimo* de arcos necesario para tener una red conexa y el *máximo* número posible para que no haya ciclos no dirigidos.

En la figura 9.3 se muestran los cinco nodos y algunos de los arcos de la figura 9.2 para ilustrar este proceso de hacer crecer un árbol mediante la colocación de un arco (rama) a la vez, hasta que se obtiene un árbol de expansión. En cada etapa del proceso existen varias alternativas para el nuevo arco, por lo que la figura 9.3 muestra sólo una de las muchas formas de construir un árbol de expansión en este caso. Sin embargo, observe cómo cada nuevo arco que se agrega satisface las condiciones especificadas en el párrafo anterior. Los árboles de expansión se estudiarán más a fondo en la sección 9.4.

FIGURA 9.3
Ejemplo de hacer crecer un árbol poniendo un arco a la vez en la red de la figura 9.2. *a*) Los nodos sin arcos; *b*) árbol con un arco; *c*) árbol con dos arcos; *d*) árbol con tres arcos; *e*) árbol de expansión.



Los árboles de expansión tienen un papel clave en el análisis de muchas redes. Por ejemplo, forman la base del *problema del árbol de mínima expansión* que se presenta en la sección 9.4. Otro ejemplo es que los árboles de expansión (factibles) corresponden a las soluciones BF del *método símplex de redes* que se analiza en la sección 9.7.

Por último, será necesario introducir terminología adicional sobre los *flujos* en redes. La cantidad máxima de flujo (quizás infinito) que puede circular en un arco dirigido se conoce como **capacidad del arco**. Entre los nodos se pueden distinguir aquellos que son generadores netos de flujo, absorbedores netos de flujo o ninguno de los dos. Un **nodo fuente** —o nodo origen— tiene la propiedad de que el flujo que *sale* del nodo supera al que *entra* a él. El caso inverso es un **nodo demanda** (o nodo destino), donde el flujo que *llega* excede al que *sale* de él. Un **nodo de trasbordo** (o intermedio) satisface la *conservación del flujo*, es decir, el flujo que entra es igual al que sale.

9.3 PROBLEMA DE LA RUTA MÁS CORTA

Aunque al final de la sección se mencionan otras versiones del problema de la ruta más corta —incluso algunas para redes dirigidas—, la atención se centrará en la siguiente versión sencilla. Considere una red *conexa* y *no dirigida* con dos nodos especiales llamados *origen* y *destino*. A cada *ligadura* (arco no dirigido) se asocia una *distancia* no negativa. El objetivo es encontrar la ruta más corta —la trayectoria con la mínima distancia total— del origen al destino.

Se dispone de un algoritmo relativamente sencillo para manejar este problema. La esencia del procedimiento es que analiza toda la red a partir del origen; identifica de manera sucesiva la ruta más corta a cada uno de los nodos en orden ascendente de sus distancias (más cortas), desde el origen; el problema queda resuelto en el momento de llegar al nodo destino. Primero se describirá el método y después se ejemplificará con la solución del problema de la ruta más corta que enfrenta la administración de Seervada Park en la sección 9.1.

Algoritmo de la ruta más corta

Objetivo de la n*-ésima iteración:* encontrar el n-ésimo nodo más cercano al origen. (Este paso se repetirá para $n = 1, 2, \ldots$ hasta que el n-ésimo nodo más cercano sea el nodo destino.)

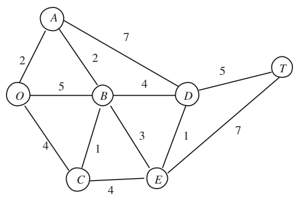
Datos de la n*-ésima iteración:* $n - 1$ nodos más cercanos al origen —que se encontró en las iteraciones previas—, incluida su ruta más corta y la distancia desde el origen. (Estos nodos y el origen se llaman *nodos resueltos*; el resto son nodos *no resueltos*.)

Candidatos para n*-ésimo nodo más cercano:* cada nodo resuelto que tiene conexión directa por una ligadura con uno o más nodos no resueltos proporciona *un* candidato, esto es, el nodo no resuelto que tiene la ligadura más *corta*. (Los empates proporcionan candidatos adicionales.)

Cálculo del n*-ésimo nodo más cercano:* para cada nodo resuelto y sus candidatos, se suma la distancia entre ellos y la distancia de la ruta más corta desde el origen a este nodo resuelto. El candidato con la distancia total más pequeña es el n-ésimo nodo más cercano —los empates proporcionan nodos resueltos adicionales—, y su ruta más corta es la que genera esta distancia.

Aplicación de este algoritmo al problema de la ruta más corta de Seervada Park

La administración de Seervada Park necesita encontrar la ruta más corta desde la entrada del parque (nodo O) hasta el mirador (nodo T) a través del sistema de caminos que se presenta en la figura 9.1. En la tabla 9.2 se encuentran los resultados que se obtuvieron al aplicar el algoritmo anterior, donde el empate del segundo nodo más cercano permite pasar directo a buscar el cuarto nodo más cercano. La primera columna (n) indica el número de la iteración. La segunda proporciona una lista de los *nodos resueltos* para comenzar la iteración actual, después de quitar los que no sirven (los



■ **TABLA 9.2** Aplicación del algoritmo de la ruta más corta al problema de Seervada Park

n	Nodos resueltos conectados directamente a nodos no resueltos	Nodo no resuelto más cercano conectado	Distancia total involucrada	*n*-ésimo nodo más cercano	Distancia mínima	Última conexión
1	*O*	*A*	2	*A*	2	*OA*
2, 3	*O* *A*	*C* *B*	4 2 + 2 = 4	*C* *B*	4 4	*OC* *AB*
4	*A* *B* *C*	*D* *E* *E*	2 + 7 = 9 4 + 3 = 7 4 + 4 = 8	*E*	7	*BE*
5	*A* *B* *E*	*D* *D* *D*	2 + 7 = 9 4 + 4 = 8 7 + 1 = 8	*D* *D*	8 8	*BD* *ED*
6	*D* *E*	*T* *T*	8 + 5 = 13 7 + 7 = 14	*T*	13	*DT*

que no tienen conexión directa con nodos no resueltos). La tercera columna da los *candidatos* para el *n*-ésimo nodo más cercano (nodos no resueltos con la ligadura *más corta* al nodo resuelto). La cuarta columna calcula la distancia de la ruta más corta desde el origen a cada candidato, esto es, la distancia al nodo resuelto más la distancia de la ligadura que va al candidato. El candidato con la suma de distancias más pequeña es el *n*-ésimo nodo más cercano al origen, según se indica en la quinta columna. Las dos últimas columnas resumen la información de este *último nodo resuelto* necesaria para pasar a las iteraciones siguientes, es decir, la distancia de la ruta más corta del origen a este nodo y la última rama en esta ruta.

Ahora se deben relacionar las columnas con la descripción del algoritmo. Los *datos para la* n-*ésima iteración* se encuentran en las columnas 5 y 6 de las iteraciones anteriores, donde los nodos resueltos de la quinta columna se enumeran después en la segunda para la iteración actual después de eliminar los que no tienen conexión directa con nodos no resueltos. Los *candidatos para el* n-*ésimo nodo más cercano* se enumeran en la tercera columna de la iteración actual. El *cálculo del n-ésimo nodo* más cercano se realiza en la columna 4 y los resultados se registran en las últimas tres columnas de la iteración actual.

La ruta más corta *desde el nodo destino hasta el origen* se puede rastrear hacia atrás en la última columna de la tabla 9.2, con lo que se obtiene $T \rightarrow D \rightarrow E \rightarrow B \rightarrow A \rightarrow$ O o bien $T \rightarrow D \rightarrow B \rightarrow A \rightarrow O$. Por tanto, se identificaron las dos opciones de ruta más corta desde el origen hasta el destino como $O \rightarrow A \rightarrow B \rightarrow E \rightarrow D \rightarrow T$ y $O \rightarrow A \rightarrow B \rightarrow D \rightarrow T$, con una distancia total de 13 millas en cualquiera de las dos.

Uso de Excel para formular y resolver problemas de la ruta más corta

Este algoritmo proporciona una manera en particular eficiente de resolver problemas de la ruta más corta. Sin embargo, algunos paquetes de programación matemática no lo incluyen. Con frecuencia incluyen el *método símplex para redes* descrito en la sección 9.7, que es otra buena opción para enfrentar estos problemas.

Como el problema de la ruta más corta es un tipo especial de problema de programación lineal, también se puede usar el método símplex general cuando no se dispone de mejores opciones. Aunque su eficiencia no se acerca a la de los algoritmos especializados en problemas grandes, es bastante adecuado aun para problemas de buen tamaño, mucho más grandes que el de Seervada Park. Excel, que se apoya en el método símplex general, proporciona un procedimiento conveniente para formular y resolver problemas de la ruta más corta con docenas de arcos y nodos.

La figura 9.4 muestra una formulación en hoja de cálculo adecuada del problema de la ruta más corta de Seervada Park. En lugar de usar el tipo de formulación de la sección 3.6 con renglo-

nes separados para cada restricción funcional del modelo de programación lineal, ésta aprovecha la estructura especial y enumera los nodos en la columna G y los *arcos* en las columnas B y C, al igual que la distancia (en millas) de cada arco en la columna E. Como cada ligadura en la red es un *arco no dirigido*, mientras que viajar por la ruta más corta tiene una dirección, cada ligadura se puede sustituir por un par de arcos *dirigidos* en direcciones opuestas. Así, las columnas B y C juntas enumeran ambas ligaduras casi verticales de la figura 9.1 (B–C y D–E) dos veces, una como arco hacia abajo y otra hacia arriba, pues se puede elegir cualquier dirección de la trayectoria. No obstante, las otras ligaduras sólo aparecen una vez como arcos de izquierda a derecha, puesto que es la única dirección de interés para elegir la ruta más corta del origen al destino.

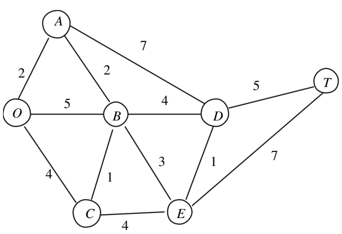
Un viaje del origen al destino se interpreta como un "flujo" de 1 por la trayectoria elegida a través de la red. Las decisiones se refieren a cuáles arcos deben incluirse en la trayectoria que se recorre. Se asigna un flujo de 1 a un arco si está incluido, mientras que el flujo es 0 si no lo está. En consecuencia, las variables de decisión son

$$x_{ij} = \begin{cases} 0 & \text{si arc } i \rightarrow j \text{ no está incluido} \\ 1 & \text{si arc } i \rightarrow j \text{ sí está incluido} \end{cases}$$

FIGURA 9.4

Formulación en hoja de cálculo del problema de la ruta más corta de Seervada Park, en la cual las celdas que cambian EnRuta (D4:D17) muestran la solución óptima que se obtiene mediante Excel Solver, y la celda objetivo DistanciaTotal(D19) que proporciona la distancia total (en millas) de esta ruta más corta. La red que se encuentra a un lado de la hoja de cálculo muestra el sistema de caminos original Seervada Park que se mostró en la figura 9.1.

	A	B	C	D	E	F	G	H	I	J
1	Problema de la ruta más corta de Seervada Park									
2										Suministro/
3		De	A	En ruta	Distancia		Nodo	Flujo		Demanda
4		O	A	1	2		O	1	=	1
5		O	B	0	5		A	0	=	0
6		O	C	0	4		B	0	=	0
7		A	B	1	2		C	0	=	0
8		A	D	0	7		D	0	=	0
9		B	C	0	1		E	0	=	0
10		B	D	0	4		T	-1	=	-1
11		B	E	1	3					
12		C	B	0	1					
13		C	E	0	4					
14		D	E	0	1					
15		D	T	1	5					
16		E	D	1	1					
17		E	T	0	7					
18										
19			Distancia total	13						



Parámetros de Solver

Celda objetivo: DistanciaTotal
Valor de la celda objetivo: ○Máximo ◉Mínimo ○

Cambio de celdas: EnRuta

Sujetas a las siguientes restricciones: FlujoNeto = SuministroDemanda

	H
3	Flujo Neto
4	=SUMASI(De,G4,EnRuta)-SUMASI(A,G4,EnRuta)
5	=SUMASI(De,G5,EnRuta)-SUMASI(A,G5,EnRuta)
6	=SUMASI(De,G6,EnRuta)-SUMASI(A,G6,EnRuta)
7	=SUMASI(De,G7,EnRuta)-SUMASI(A,G7,EnRuta)
8	=SUMASI(De,G8,EnRuta)-SUMASI(A,G8,EnRuta)
9	=SUMASI(De,G9,EnRuta)-SUMASI(A,G9,EnRuta)
10	=SUMASI(De,G10,EnRuta)-SUMASI(A,G10,EnRuta)

Opciones de Solver

☑ Adoptar modelo lineal
☑ Asumir no negativos

Nombre de rango	Celdas
Distancia	E4:E17
De	B4:B17
FlujoNeto	H4:H10
Nodos	G4:G10
EnRuta	D4:D17
SuministroDemanda	J4:J10
A	C4:C17
DistanciaTotal	D19

	C	D
19	**DistanciaTotal**	=SUMAPRODUCTO(D4:D17,E4:E17)

Recuadro de aplicación

Canadian Pacific Railway (CPR), fundada en 1881, fue la primera compañía ferrocarrilera transcontinental en Norteamérica. CPR transporta carga a través de una red de más de 14 000 millas que se extiende desde Montreal hasta Vancouver y del noroeste al medio oriente de Estados Unidos. Sus alianzas con otras compañías transportistas extienden el mercado de CPR a los principales centros de negocios de México también.

CPR recibe todos los días aproximadamente 7 000 nuevos embarques de sus clientes que viajan a destinos dentro de Norteamérica y al extranjero. CPR debe desplazar dichos embarques en carros de ferrocarril a través de la red ferroviaria, donde un determinado carro puede cambiar varias veces de una locomotora a otra antes de llegar a su destino. CPR debe coordinar los embarques de acuerdo con sus planes de operación de 1 600 locomotoras, 65 000 carros de ferrocarril, más de 5 000 miembros de la tripulación de los trenes y 250 estaciones de ferrocarril.

La gerencia de CPR contrató a una firma consultora de investigación de operaciones llamada Multimodal Applied Systems para trabajar con los empleados de CPR en el desrrollo de un método de investigación de operaciones para resolver este problema. Se utilizó gran variedad de técnicas de investigación de operaciones a fin de diseñar una nueva estrategia de operaciones. Sin embargo, las bases del método consistían en representar el flujo de bloques de carros de ferrocarril como un flujo a través de una red donde cada nodo correspondía a una ubicación y a un determinando instante de tiempo. Así, dicha representación permitía la aplicación de las técnicas de optimización de la red. Por ejemplo, se resuelve diariamente gran cantidad de problemas de trayectoria más corta como parte de este método.

Esta aplicación de la investigación de operaciones representa un ahorro para CPR de aproximadamente **100 millones de dólares** *anuales*. Se han incrementado de manera sustancial la productividad laboral, la productividad del uso de las locomotoras, el consumo de combustible y la velocidad de los carros de ferrocarril. Además, CPR brinda a sus clientes tiempos de entrega confiables, y ha recibido innumerables reconocimientos por su mejora en el servicio. Esta aplicación de las técnicas para la optimización de la red también llevó a CPR a ganar en 2003 el primer lugar en la competencia internacional por el Premio Franz Edelman al desempeño en investigación de operaciones y las ciencias de la administración.

Fuente: P. Ireland, R. Case, J. Fallis, C. Van Dyke, J. Kuehn y M. Meketon: "The Canadian Pacific Railway Transforms Operations by Using Models to Develop Its Operating Plans". *Interfaces*, **34**(1): 5-14, enero-febrero, 2004. (En nuestra página en internet www.mhhe.com/hillier se proporciona una liga hacia este artículo.)

para cada arco en consideración. Los valores de estas variables de decisión se introducen en las celdas cambiantes EnRuta (D4:D17).

Se puede pensar en cada nodo como que tiene un flujo de 1 si está en la trayectoria seleccionada y sin flujo en otro caso. El *flujo neto* generado en un nodo es el *flujo que sale* menos el *flujo que entra*, de manera que el flujo neto es 1 en el origen, –1 en el destino y 0 en el resto de los nodos. Estos requisitos de los flujos netos se especifican en la columna J de la figura 9.4. Al usar las ecuaciones en la parte inferior de la figura, cada celda de la columna H calcula el flujo neto real en ese nodo mediante la suma del flujo que sale y la resta del flujo que entra. Las restricciones correspondientes, FlujoNeto (H4:H10) = DemandaDeSuministro (J4:J10) se especifican en el cuadro de diálogo de Solver.

La celda objetivo DistanciaTotal (D19) proporciona la distancia total en millas de la trayectoria que se eligió al usar la ecuación para esta celda dada en la parte baja de la figura 9.4. El objetivo de *minimizar* esta celda se especifica en el cuadro de diálogo de Solver. La solución que se presenta en la columna D es una solución óptima que se obtiene después de oprimir el botón de resolver. Esta solución es, por supuesto, una de las dos rutas más cortas identificadas antes por el algoritmo de la ruta más corta.

Otras aplicaciones

No todas las aplicaciones del problema de la ruta más corta involucran minimizar la distancia recorrida de un origen a un destino. En realidad, es posible que ni siquiera se refieran a un viaje. Las ligaduras (o arcos) pueden representar actividades de otro tipo, por lo que escoger una trayectoria a través de la red significa seleccionar la mejor secuencia de actividades. Por ello, los números que indican las "longitudes" de las ligaduras quizá sean, por ejemplo, los costos de las actividades, en cuyo caso el objetivo sería determinar qué secuencia de actividades minimiza el costo total. En la sección Worked Examples del sitio en internet de este libro se incluye otro ejemplo de este tipo que ilustra su formulación como un problema de la ruta más corta y su solución mediante el uso de un algoritmo especial para esos problemas, o por medio del Excel Solver con una formulación en hoja de cálculo.

Las siguientes son tres categorías de aplicaciones.

1. Minimizar la *distancia* total recorrida, como en el ejemplo de Seervada Park.
2. Minimizar el *costo* total de una secuencia de actividades (como en el problema 9.3-3).
3. Minimizar el *tiempo* total de una secuencia de actividades (los problemas 9.3-6 y 9.3-7 son de este tipo).

Incluso, es posible que las tres categorías coexistan en el *mismo* problema. Por ejemplo, suponga que se desea encontrar la mejor ruta de un lugar a otro a través de cierto número de lugares intermedios. En este tipo de casos se tiene la opción de definir la mejor ruta como la que minimiza la *distancia* total recorrida, la que minimiza el *costo* total en el que se incurre o la que minimiza el *tiempo* total que se requiere. (El problema 9.3-2 ilustra esta aplicación.)

Muchas aplicaciones requieren encontrar la trayectoria *dirigida* más corta del origen al destino a través de una red *dirigida*. El algoritmo que acaba de presentarse se puede modificar con facilidad para que maneje trayectorias dirigidas en cada iteración. En particular, cuando se identifican candidatos como el n-ésimo nodo más cercano, sólo se deben considerar los arcos dirigidos *desde* un nodo resuelto *hacia* un nodo no resuelto.

Otra versión del problema de la ruta más corta es encontrar las rutas más cortas del origen a *todos* los demás nodos de la red. Observe que el algoritmo obtiene las rutas más cortas a cada nodo que está más cerca del origen que del destino. Entonces, si todos los nodos son destinos potenciales, la única modificación que se necesita es que el algoritmo no se detenga hasta que todos los nodos se hayan resuelto.

Una versión aún más general del problema de la ruta más corta es encontrar la ruta más corta desde *cada* nodo a todos los demás. Otra opción es eliminar la restricción de que las "distancias" —valores de los arcos— sean no negativas. Se pueden poner también restricciones sobre las trayectorias posibles. En ocasiones, todas estas variaciones surgen en la práctica, razón por la cual han sido estudiadas por los investigadores.

Los algoritmos de una gran variedad de problemas de optimización de análisis combinatorio, como los problemas de diseño de rutas de vehículos, con frecuencia utilizan como parte de sus subrutinas la solución de un gran número de problemas de la ruta más corta. Aunque no se dispone de espacio suficiente para profundizar en este tema, tal vez esta aplicación sea una de las más importantes de este problema.

9.4 PROBLEMA DEL ÁRBOL DE EXPANSIÓN MÍNIMA

El problema del árbol de expansión mínima tiene algunas similitudes con la versión principal del problema de la ruta más corta que se presentó en la sección anterior. En ambos casos se considera una red *no dirigida* y *conexa*, en la que la información dada incluye alguna medida de *longitud* positiva —distancia, costo, tiempo, etc.— asociada con cada ligadura. Los dos problemas involucran también el hecho de seleccionar un conjunto de ligaduras con la *longitud total más corta* entre todos los conjuntos de ligaduras que satisfacen cierta propiedad. En el caso del problema de la ruta más corta, esta propiedad es que la ligadura seleccionada debe proporcionar una trayectoria entre el origen y el destino. Para el árbol de expansión mínima la propiedad que se requiere es que las ligaduras seleccionadas deben proporcionar una trayectoria entre *cada* par de nodos.

El problema del árbol de expansión mínima se puede resumir de la siguiente manera:

1. Se tienen los *nodos* de una red pero no las *ligaduras*. En su lugar se proporcionan las ligaduras potenciales y la longitud positiva de cada una si se insertan en la red. (Las medidas alternativas para la longitud de una ligadura incluyen distancia, costo y tiempo.)
2. Se desea diseñar la red con suficientes ligaduras para satisfacer el requisito de que haya un camino entre *cada* par de nodos.
3. El objetivo es satisfacer este requisito de manera que se minimice la longitud total de las ligaduras insertadas en la red.

Una red con n nodos requiere de sólo $(n - 1)$ ligaduras para proporcionar una trayectoria entre cada par de nodos. No deben usarse más ligaduras puesto que ello aumentaría, sin necesidad, la longitud total de las ligaduras seleccionadas. Las $(n - 1)$ ligaduras deben elegirse de tal manera

que la red resultante —con sólo las ligaduras seleccionadas— forme un *árbol de expansión*, según la definición que se presentó en la sección 9.2. Por lo tanto, el problema es encontrar el árbol de expansión con la longitud total mínima de sus ligaduras.

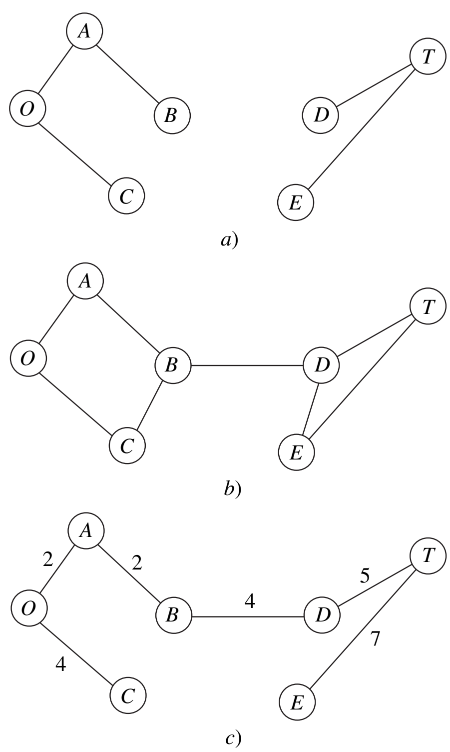
La figura 9.5 ilustra el concepto de árbol de expansión del problema de Seervada Park (sección 9.1). La figura 9.5*a no* es un árbol de expansión, pues los nodos *O*, *A*, *B* y *C* no están conectados con los nodos *D*, *E* y *T*. Se necesita una ligadura más para hacer esta conexión. En realidad, esta red consta de dos árboles, uno para cada uno de estos dos conjuntos de nodos. Las ligaduras de la figura 9.5*b* sí se *expanden* por toda la red —es decir, es una gráfica conexa según la definición de la sección 9.2—, pero *no* es un árbol porque tiene dos *ciclos* (*O–A–B–C–O* y *D–T–E–D*), esto es, tiene demasiadas ligaduras. Como el problema de Seervada Park tiene $n = 7$ nodos, en la sección 9.2 se indicó que una red debe tener exactamente $n - 1 = 6$ ligaduras y *ningún ciclo* para calificar como árbol de expansión. Esta condición se logra en la figura 9.5*c*, por lo que esta red es una solución *factible* —con una longitud total de 24 millas en las ramas o ligaduras— para el problema del árbol de expansión mínima. (Se verá que esta solución no es *óptima*, puesto que es posible construir un árbol de expansión con sólo 14 millas en sus ramas.)

Algunas aplicaciones

A continuación se proporciona una lista de algunos tipos importantes de aplicaciones de este problema.

1. Diseño de redes de telecomunicación (redes de fibra óptica, de computadoras, telefónicas, de televisión por cable, etcétera).
2. Diseño de redes de transporte para minimizar el costo total de proporcionar las ligaduras (vías ferroviarias, carreteras, etcétera).
3. Diseño de una red de líneas de transmisión de energía eléctrica de alto voltaje.
4. Diseño de una red de cableado de equipo eléctrico —como sistemas de cómputo— para minimizar la longitud total de cable.
5. Diseño de una red de tuberías para conectar varias localidades.

FIGURA 9.5
Ilustración del concepto de árbol de expansión mínima del problema de Seervada Park: *a*) no es un árbol de expansión; *b*) no es un árbol de expansión; *c*) árbol de expansión.



En esta era de la supercarretera de la información, las aplicaciones del primer tipo han cobrado una importancia especial, pues en una red de telecomunicaciones sólo es necesario insertar suficientes ligaduras para que proporcionen una trayectoria entre cada par de nodos, de modo que el diseño de tales redes es una aplicación clásica del problema del árbol de expansión mínima. Debido a que en la actualidad algunas redes de comunicación cuestan muchos millones de dólares, es muy importante optimizar su diseño al encontrar el árbol de expansión mínima.

Un algoritmo

El problema del árbol de expansión mínima se puede resolver de una forma bastante directa, puesto que se trata de uno de los pocos problemas de IO en el que ser *codicioso* en cada etapa del procedimiento de solución conduce al final a una solución óptima. Así, con el inicio en cualquier nodo, la primera etapa consiste en elegir la rama más corta posible a otro nodo, sin preocuparse del efecto que esta elección pueda tener en las decisiones posteriores. En la segunda etapa se trata de identificar el nodo no conectado que esté más cerca de cualquiera de los dos que se acaban de conectar y después agregar la ligadura correspondiente a la red. Este proceso se repite, según el resumen que se presenta a continuación, hasta conectar todos los nodos. (Observe que éste es el proceso que se ilustró en la figura 9.3 para construir un árbol de expansión, pero ahora con la regla específica para seleccionar cada ligadura nueva.) Se garantiza que la red resultante es un árbol de expansión mínima.

Algoritmo del problema del árbol de expansión mínima

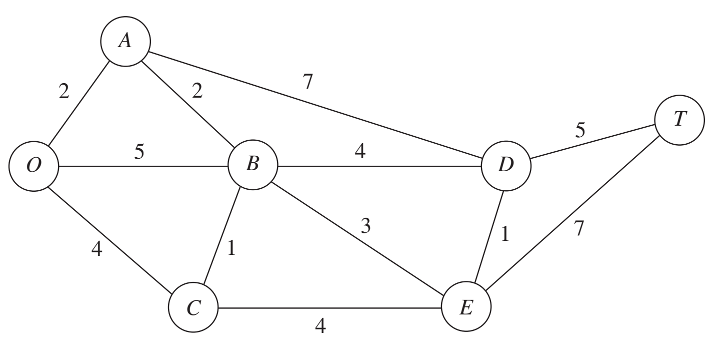
1. Se selecciona, de manera arbitraria, cualquier nodo y se conecta, es decir, se agrega una ligadura al nodo distinto más cercano.
2. Se identifica el nodo no conectado más cercano a un nodo conectado y se conectan estos dos nodos, esto es, se agrega una ligadura entre ellos. Este paso se repite hasta que todos los nodos están conectados.
3. Rompimiento de empates: los empates del nodo más cercano distinto (paso 1) o del nodo no conectado más cercano (paso 2), se pueden romper en forma arbitraria, pero el algoritmo debe llegar a una solución óptima. No obstante, estos empates son señal de que pueden existir (pero no necesariamente) soluciones óptimas múltiples. Todas esas soluciones se pueden identificar si se trabaja con las demás formas de romper los empates al final.

La manera más rápida de ejecutar este algoritmo en forma manual es el enfoque gráfico que se ilustra en seguida.

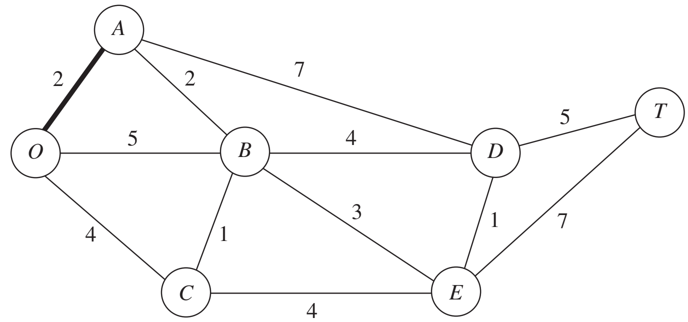
Aplicación de este algoritmo al problema del árbol de expansión mínima de Seervada Park

La administración de Seervada Park (sección 9.1) debe determinar los caminos bajo los cuales se deben tender las líneas telefónicas para conectar todas las estaciones con una longitud total mínima de cable. Se describirá paso a paso la solución de este problema con base en los datos que se proporcionan en la figura 9.1.

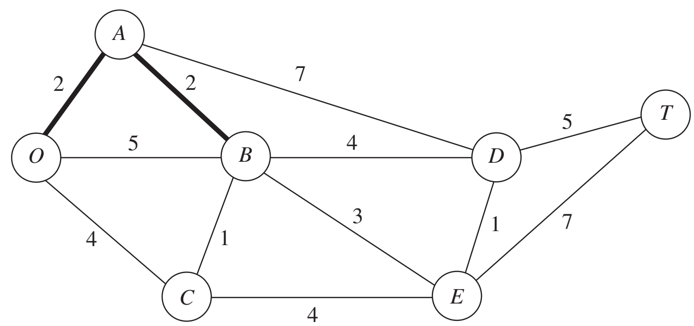
Los nodos y distancias del problema se resumen a continuación, en donde las líneas delgadas ahora representan ligaduras *potenciales*.



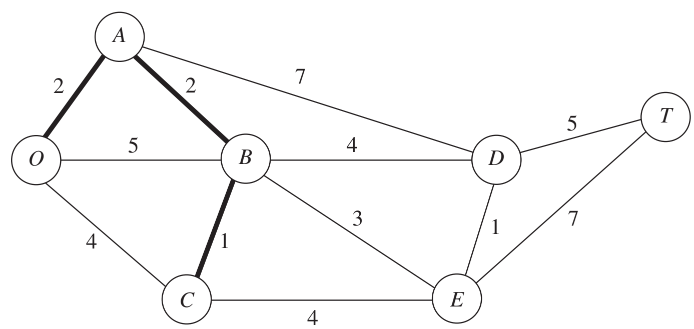
En forma arbitraria, se selecciona el nodo *O* como inicio. El nodo no conectado más cercano a *O* es *A*. Se conecta el nodo *A* con el nodo *O*.



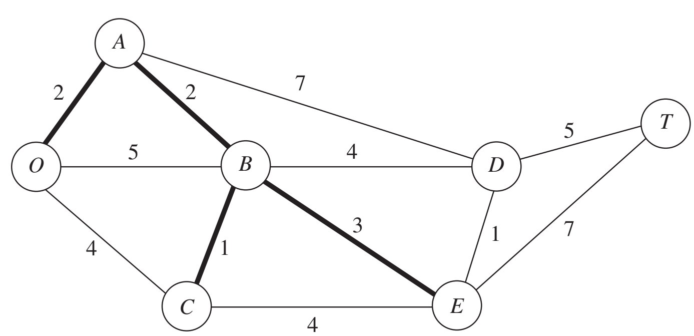
El nodo no conectado más cercano a cualesquiera de los nodos *O* o *A* es el nodo *B* (más cercano a *A*). Se conecta el nodo *B* con el nodo *A*.



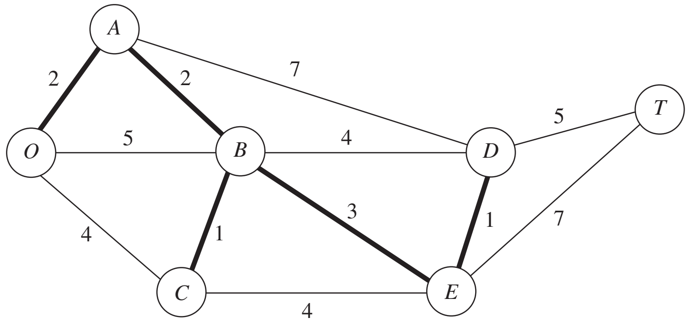
El nodo no conectado más cercano a *O*, *A* o *B* es el nodo *C* (más cercano a *B*). Se conecta el nodo *C* con el nodo *B*.



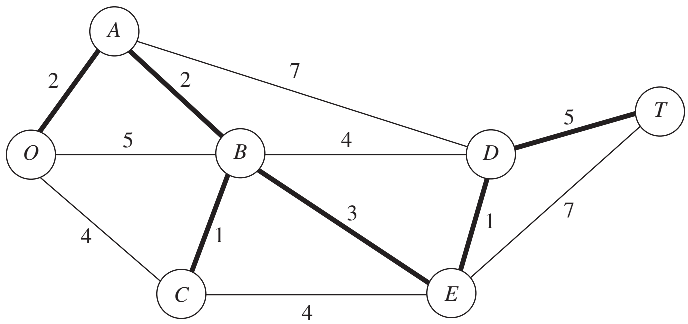
El nodo no conectado más cercano a *O*, *A*, *B* o *C* es el nodo *E* (más cercano a *B*). Se conecta el nodo *E* con el nodo *B*.



El nodo no conectado más cercano a los nodos *O*, *A*, *B*, *C* o *E* es el nodo *D* (más cercano a *E*). Se conecta el nodo *D* con el nodo *E*.



El único nodo no conectado es el nodo *T*. Está más cerca del nodo *D*. Se conecta el nodo *T* con el nodo *D*.



Todos los nodos han quedado conectados, por lo que ésta es la solución (óptima) que se buscaba. La longitud total de las ramas es de 14 millas.

Aunque con este procedimiento a primera vista puede parecer que la elección del nodo inicial afectará la solución final —y la longitud total de las ligaduras—, en realidad no es así. Se sugiere que verifique este hecho en el caso del ejemplo, mediante otra aplicación del algoritmo, pero con un nodo inicial distinto de *O*.

Se considera que dentro de este capítulo el problema del árbol de expansión mínima es el que cae dentro de la amplia categoría de diseño de redes. En esta categoría, el objetivo es diseñar la red más apropiada para el problema dado —con frecuencia se trata de sistemas de transporte— y no de analizar una red ya diseñada. La referencia seleccionada 6 proporciona una investigación en esta importante área.

9.5 PROBLEMA DE FLUJO MÁXIMO

Recuerde que el tercer problema al que se enfrenta la administración de Seervada Park (vea la sección 9.1) durante la temporada pico es determinar las rutas de algunos viajes de tranvía desde la entrada del parque (estación *O* en la figura 9.1) hasta el mirador (estación *T*), de manera que el número de viajes diarios sea máximo. (Cada tranvía debe regresar por la misma ruta que tomó de ida, por lo que el análisis se hará sólo sobre los viajes de ida.) Para evitar perturbaciones innecesarias a la ecología y a la vida silvestre se impusieron límites superiores estrictos sobre el número de viajes de salida permitidos hacia el mirador para cada camino individual en la dirección de ida. Para cada camino, la dirección del viaje de ida se indica mediante una flecha en la figura 9.6. El número que aparece en la base de la flecha proporciona el límite superior de viajes en la dirección de salida de la estación. Dados los límites, una *solución factible* es enviar 7 tranvías al día, 5 por la ruta $O \to B \to E \to T$, 1 por la ruta $O \to B \to C \to E \to T$ y 1 por la ruta $O \to B \to C \to E \to D \to T$. Esta solución bloquea el uso de cualquier ruta que comience con $O \to C$, debido a que las capacidades de $E \to T$ y $E \to D$ están saturadas. Es sencillo encontrar mejores soluciones factibles. Es necesario

considerar muchas *combinaciones* de rutas —y el número de viajes asignados a cada una— para encontrar la(s) ruta(s) que maximice(n) el número de viajes al día. Este tipo de problemas se conoce como *problema del flujo máximo*.

En términos generales, el problema de flujo máximo se puede describir de la siguiente manera.

1. Todo flujo a través de una red conexa dirigida se origina en un nodo, llamado **origen**, y termina en otro nodo llamado **destino** (el origen y el destino en Seervada Park son la entrada al parque en el nodo *O* y el mirador en el nodo *T*, respectivamente).
2. Los nodos restantes son *nodos de trasbordo* (en el problema de Seervada Park son los nodos *A*, *B*, *C*, *D* y *E*).
3. Se permite el flujo a través de un arco sólo en la dirección indicada por la flecha, donde la cantidad máxima de flujo está dada por la *capacidad* del arco. En el origen, todos los arcos señalan hacia afuera. En el *destino*, todos señalan hacia el nodo.
4. El objetivo es maximizar la cantidad total de flujo del origen al destino. Esta cantidad se mide en cualquiera de las dos maneras equivalentes, esto es, la cantidad que *sale del origen* o la cantidad que *entra al destino*.

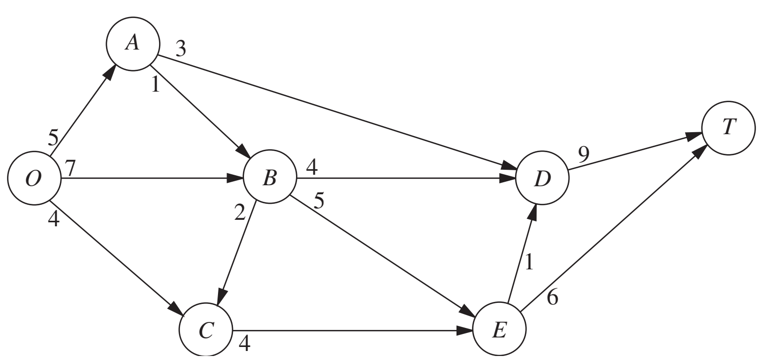
Algunas aplicaciones

A continuación se mencionan algunos tipos de aplicaciones comunes del problema del flujo máximo.

1. Maximizar el flujo a través de la red de distribución de una compañía desde sus fábricas hasta sus clientes.
2. Maximizar el flujo a través de la red de suministros de una compañía de proveedores a las fábricas.
3. Maximizar el flujo de petróleo por un sistema de tuberías.
4. Maximizar el flujo de agua a través de un sistema de acueductos.
5. Maximizar el flujo de vehículos por una red de transporte.

En algunas de estas aplicaciones, el flujo a través de la red se puede originar en más de un nodo y también puede terminar en más de uno, aunque el problema de flujo máximo puede tener sólo un origen y un destino. Por ejemplo, una red de distribución de una compañía tiene varias fábricas y múltiples clientes. En este caso se recurre a una reformulación ingeniosa para ajustar esta situación al problema de flujo máximo. Se trata de aumentar la red original para que incluya *un origen ficticio*, *un destino ficticio* y algunos arcos nuevos. El origen ficticio se maneja como el nodo que da origen a todo el flujo que en realidad se origina en algunos otros nodos. En cada uno de estos otros nodos se inserta un nuevo arco que va desde el origen ficticio hasta este nodo, donde la capacidad del arco es igual al flujo máximo que se puede originar en este nodo. De manera similar, el destino ficticio se trata como el nodo que absorbe todo el flujo que, en realidad, termina en algún otro nodo. Por lo tanto, se coloca un nuevo arco desde cada uno de los otros nodos hasta el destino ficticio con capacidad igual al flujo máximo que en realidad termina en este nodo. Debido

FIGURA 9.6
Problema del flujo máximo de Seervada Park.



a estos cambios, todos los nodos de la red original se convierten en nodos de trasbordo para que la red aumentada tenga un solo origen (la fuente ficticia) y un solo destino (el destino ficticio) y se ajuste al problema del flujo máximo.

Un algoritmo

Como el problema del flujo máximo se puede formular como un *problema de programación lineal* (vea el problema 9.5-2), se puede resolver con el método símplex, así como usar cualquier software como los que se presentaron en los capítulos 3 y 4. Sin embargo, se dispone de un *algoritmo de trayectorias aumentadas* mucho más eficiente. Este algoritmo se basa en dos conceptos intuitivos, el de una *red residual* y el de una *trayectoria aumentada*.

Una vez que se han asignado flujos a los arcos de la red original, la **red residual** muestra las capacidades *restantes* —llamadas **capacidades residuales**— para asignar flujos *adicionales*. Por ejemplo, considere el arco $O \to B$ de la figura 9.6, que tiene una capacidad de 7. Ahora suponga que los flujos asignados incluyen un flujo de 5 a través de este arco, lo que deja una capacidad residual de $7 - 5 = 2$ para cualquier asignación de flujo adicional a través de $O \to B$. Este estado se describe en la red residual de la siguiente manera.



El número sobre el arco junto a un nodo señala la capacidad residual del flujo *desde* ese nodo *hasta* el otro. Por lo tanto, además de la capacidad residual de 2 del flujo de O a B, el 5 de la derecha indica una capacidad residual de 5 para asignar un flujo desde B hasta O, es decir, para cancelar algún flujo asignado antes de O a B.

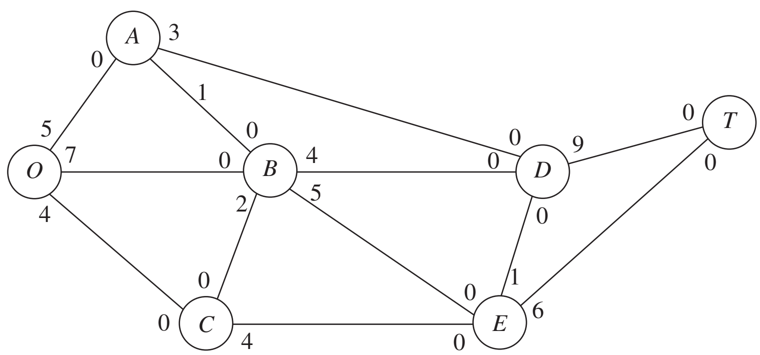
De inicio, antes de asignar cualquier flujo, la red residual tiene la apariencia que se muestra en la figura 9.7. Todos los arcos de la red original (figura 9.6) se cambiaron de un *arco dirigido* a un arco *no dirigido*. No obstante, las capacidades en la dirección original son las mismas y las capacidades en la dirección opuesta son cero, de manera que las restricciones sobre los flujos no cambian.

Después, siempre que se asigna una cantidad de flujo a un arco, esa cantidad se *resta* de la capacidad residual en la misma dirección y se *suma* a la capacidad residual en la dirección opuesta.

Una **trayectoria de aumento** es una trayectoria dirigida del nodo origen al nodo destino en la red residual, tal que *todos* los arcos en esta trayectoria tienen capacidad residual *estrictamente positiva*. El *mínimo* de estas capacidades residuales se llama *capacidad residual de la trayectoria de aumento* porque representa la cantidad de flujo que es factible agregar en toda la trayectoria. Por lo tanto, cada trayectoria de aumento proporciona una oportunidad de aumentar más el flujo a través de la red original.

El algoritmo de la trayectoria de aumento selecciona varias veces una trayectoria de aumento y agrega un flujo igual a su capacidad residual a la trayectoria en la red original. Este proceso continúa hasta que no hay trayectorias de aumento, con lo que el flujo del nodo fuente al nodo destino no

■ **FIGURA 9.7**
Red residual inicial del problema de flujo máximo de Seervada Park.



puede crecer. La clave para asegurar que la solución final es óptima por necesidad es el hecho de que las trayectorias de aumento pueden cancelar flujos asignados con anterioridad en la red original; así, una selección indiscriminada de trayectorias para asignar flujos no puede evitar el uso de una combinación mejor de asignaciones de flujos.

Para resumir, cada *iteración* del algoritmo consiste en los tres pasos siguientes.

Algoritmo de la trayectoria de aumento del problema de flujo máximo[1]

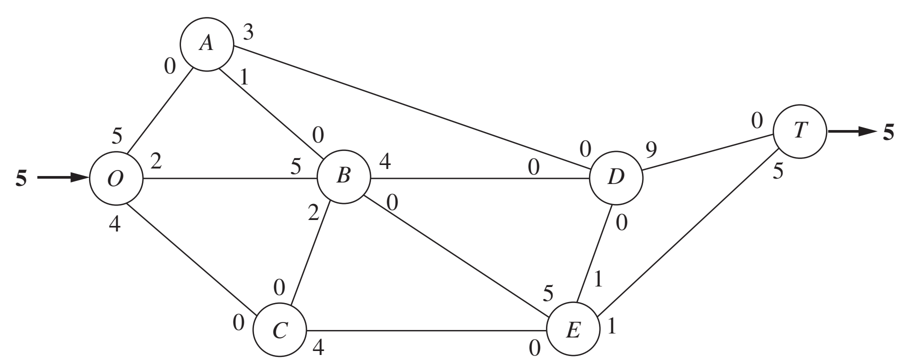
1. Se identifica una trayectoria de aumento cuando se encuentra alguna trayectoria dirigida del origen al destino en la red residual, tal que cada arco sobre ella tenga capacidad residual estrictamente positiva. (Si no existe una, los flujos netos asignados constituyen un patrón de flujo óptimo.)
2. Cuando se encuentra el *mínimo* de las capacidades residuales de los arcos sobre esta trayectoria se identifica la capacidad residual c^* de esta trayectoria de aumento. Se *aumenta* en c^* el flujo de esta trayectoria.
3. Se *disminuye* en c^* la capacidad residual de cada arco en esta trayectoria de aumento. Se aumenta en c^* la capacidad residual de cada arco en la dirección opuesta en esta trayectoria. Se regresa al paso 1.

Cuando se lleva a cabo el paso 1, con frecuencia habrá varias alternativas de trayectorias de aumento entre las cuales se podrá escoger. Aunque la estrategia algorítmica para elegir es importante para elevar la eficiencia de las aplicaciones a gran escala, no se profundizará en este tema relativamente especializado. (Más adelante, en esta misma sección, se describe un procedimiento sistemático para encontrar una trayectoria aumentada.) En consecuencia, en el siguiente ejemplo —y los problemas del final del capítulo—, la selección se hará en forma arbitraria.

Aplicación del algoritmo al problema de flujo máximo de Seervada Park

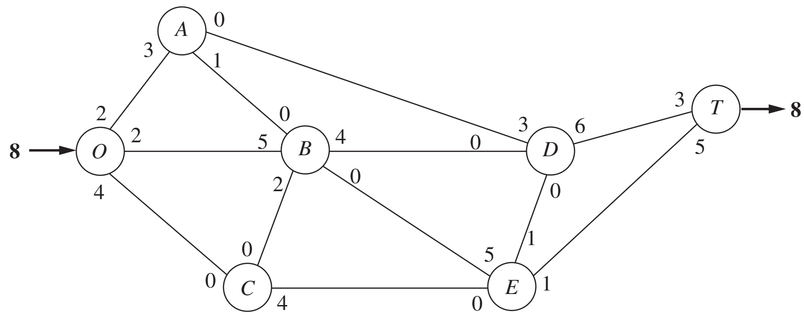
La aplicación de este algoritmo al problema de Seervada Park —vea la red original en la figura 9.6— conduce a los siguientes resultados. (También vea la sección Worked Examples del sitio de internet de este libro donde se incluye un **segundo ejemplo** de la aplicación de este algoritmo.) A partir de la red residual inicial en la figura 9.7, se proporciona la nueva red residual después de una o dos iteraciones, donde la cantidad total de flujo de O a T que se logró hasta el momento se muestra en **negritas** (junto a los nodos O y T).

Iteración 1: en la figura 9.7, una de las trayectorias de aumento es $O \to B \to E \to T$ que tiene capacidad residual igual al min $\{7, 5, 6\} = 5$. Si se asigna un flujo de 5 a esta trayectoria, la red residual que resulta es



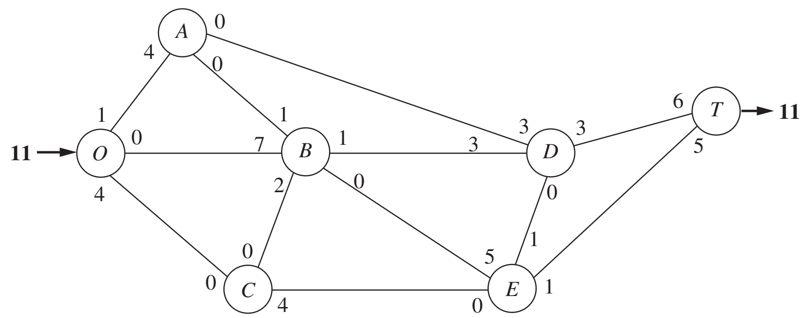
[1] Se supone que las capacidades de los arcos son números enteros o racionales.

Iteración 2: se asigna un flujo de 3 a la trayectoria de aumento $O \rightarrow A \rightarrow D \rightarrow T$. La red residual que resulta es



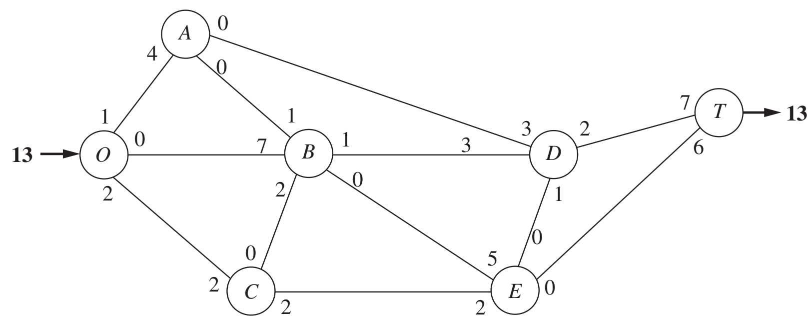
Iteración 3: se asigna un flujo de 1 a la trayectoria de aumento $O \rightarrow A \rightarrow B \rightarrow D \rightarrow T$.

Iteración 4: se asigna un flujo de 2 a la trayectoria de aumento $O \rightarrow B \rightarrow D \rightarrow T$. La red residual que resulta es

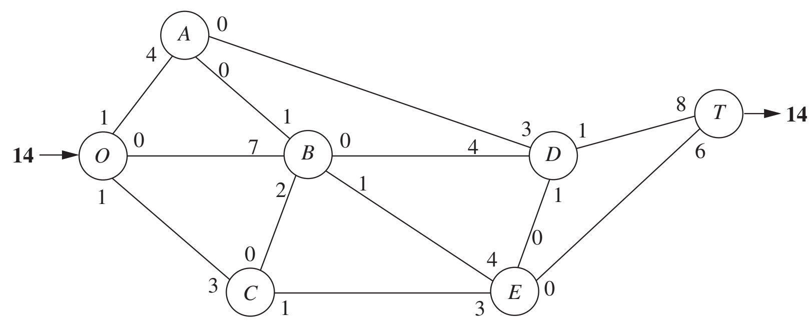


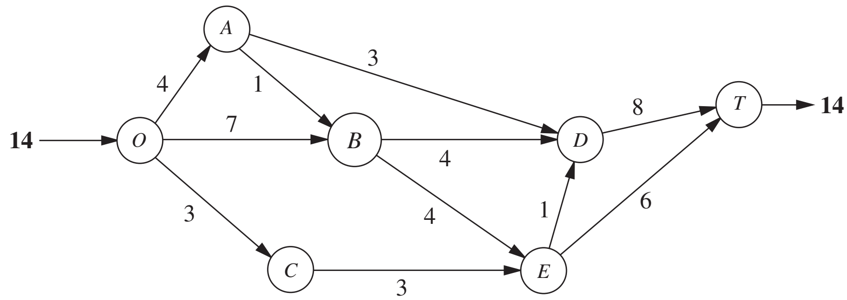
Iteración 5: se asigna un flujo de 1 a la trayectoria de aumento $O \rightarrow C \rightarrow E \rightarrow D \rightarrow T$.

Iteración 6: se asigna un flujo de 1 a la trayectoria de aumento $O \rightarrow C \rightarrow E \rightarrow T$. La red residual resultante es



Iteración 7: se asigna un flujo de 1 a la trayectoria de aumento $O \rightarrow C \rightarrow E \rightarrow B \rightarrow D \rightarrow T$. La red residual que resulta es





■ **FIGURA 9.8**
Solución óptima del problema de flujo máximo de Seervada Park.

Ya no existen trayectorias de aumento, por lo que el patrón de flujo actual es óptimo.

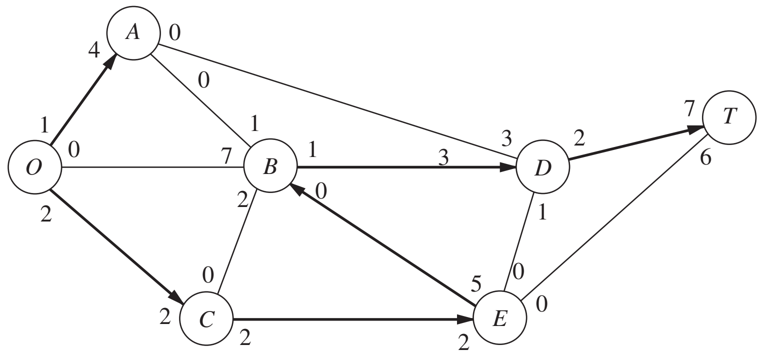
El patrón de flujo actual se puede identificar ya sea por medio de la acumulación de las asignaciones de flujo o mediante la comparación de las capacidades residuales finales con las capacidades originales de los arcos. Si se emplea este método, existe un flujo a través de un arco si la capacidad residual final es menor que la capacidad original. La magnitud de este flujo es igual a la diferencia entre estas capacidades. Al aplicar este método de comparación de la red residual que se obtuvo en la última iteración ya sea en la figura 9.6 o en la 9.7, se obtiene el patrón de flujo óptimo que se muestra en la figura 9.8.

Este ejemplo ilustra en forma sencilla la razón para sustituir cada arco dirigido $i \rightarrow j$ de la red original por un arco no dirigido en la red residual y después aumentar c^* unidades a la capacidad residual de $j \rightarrow i$ cuando se asigna un flujo de c^* al arco $i \rightarrow j$. Sin este refinamiento, las primeras seis iteraciones no cambian, pero en ese momento parecería que ya no quedan trayectorias de aumento ya que la capacidad de flujo real, sin usar de $E \rightarrow B$, es cero. El refinamiento permite agregar la asignación de un flujo de l a $O \rightarrow C \rightarrow E \rightarrow B \rightarrow D \rightarrow T$ en la iteración 7. Esta asignación adicional cancela el flujo de 1 asignado en la iteración 1 ($O \rightarrow B \rightarrow E \rightarrow T$) y lo sustituye por las asignaciones de una unidad a las dos rutas $O \rightarrow B \rightarrow D \rightarrow T$ y $O \rightarrow C \rightarrow E \rightarrow T$.

Búsqueda de una trayectoria de aumento

La parte más difícil de este algoritmo, cuando se trabaja con *redes grandes*, es encontrar una trayectoria de aumento. Esta tarea se puede simplificar con un procedimiento sistemático. Se comienza por determinar todos los nodos que se pueden alcanzar desde el origen con un solo arco con capacidad residual estrictamente positiva. Después, en el caso de cada uno de estos nodos alcanzados, se determinan todos los *nuevos* nodos —entre los que no han sido alcanzados— a los que se puede llegar desde este nodo con un solo arco con capacidad residual estrictamente positiva. Este procedimiento se repite con los nuevos nodos a medida que se llega a ellos. El resultado será la identificación de un árbol con todos los nodos a los que se puede llegar desde el origen, a lo largo de una trayectoria con capacidad de flujo residual estrictamente positiva. Este *procedimiento de abanico* siempre identificará una trayectoria de aumento, si existe. En la figura 9.9 se ilustra este proceso para el caso de la red residual que se obtuvo en la *iteración 6* del ejemplo anterior.

Aunque el procedimiento de la figura 9.9 es relativamente directo, es útil poder reconocer cuándo se tiene un patrón óptimo sin tener que buscar de manera exhaustiva una ruta que no existe. A veces es posible reconocer esto con el resultado de un teorema importante de teoría de redes, conocido como *teorema del flujo-máximo corte-mínimo*. Un **corte** se puede definir como cualquier conjunto de arcos dirigidos que contienen al menos un arco de cada trayectoria dirigida que va del nodo origen al nodo destino. En general, hay muchas formas de dividir una red para formar un corte que ayude a analizarla. El **valor del corte** es la suma de las capacidades de los arcos (en la dirección especificada) del corte. El **teorema de flujo-máximo corte-mínimo** establece que para cualquier red con un solo nodo origen y un solo nodo destino, el *flujo máximo factible* del origen al destino es *igual al valor del corte mínimo* de todos los cortes de la red. Así, si F denota la cantidad de flujo desde el origen hasta el destino de cualquier patrón de flujo factible, el valor de cualquier corte proporciona una cota superior de F, y el menor de los valores de los cortes es igual al máximo valor de F. Entonces, si se puede encontrar, en la red original, un corte cuyo valor sea igual al



■ **FIGURA 9.9**
Procedimiento para encontrar una trayectoria de aumento en la iteración 7 del problema de flujo máximo de Seervada Park.

valor actual de *F* que se encontró con el procedimiento de solución, el patrón de flujo actual debe ser *óptimo*. Finalmente se alcanza la optimalidad siempre que exista un corte cuyo valor sea cero en la red residual.

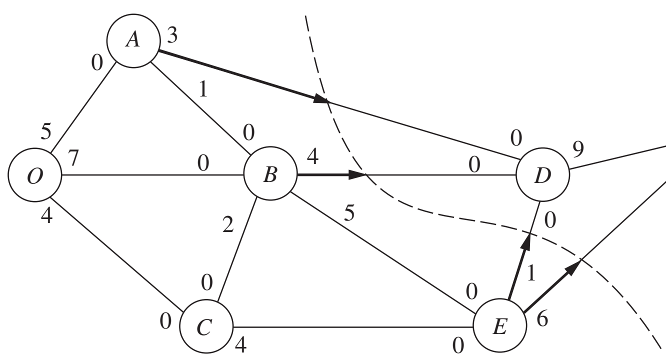
Para ilustrar lo que decimos, considere la red de la figura 9.7. Un corte interesante se muestra en la figura 9.10. Observe que el valor del corte es $3 + 4 + 1 + 6 = 14$ que, según se había encontrado, corresponde al máximo valor de *F*, por lo que se trata de un corte mínimo. Observe también que en la red residual que se obtuvo en la iteración 7, en donde $F = 14$, el corte correspondiente tiene valor cero. Si este detalle se hubiera observado, no habría sido necesario buscar más trayectorias de aumento.

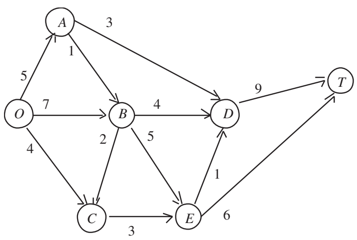
Uso de Excel para formular y resolver problemas de flujo máximo

La mayoría de los problemas de flujo máximo que surgen en la práctica son muy grandes y en ocasiones mucho más grandes que el problema de Seervada Park. Algunos de ellos tienen miles de nodos y arcos. El algoritmo de la trayectoria aumentada es mucho más eficiente que el método símplex general para resolver esos problemas grandes. Sin embargo, para problemas de tamaño modesto, una alternativa razonable y conveniente es usar Excel y el Solver basado en el método símplex general.

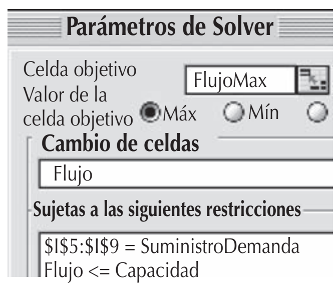
En la figura 9.11 se muestra una formulación en hoja de cálculo del problema de flujo máximo de Seervada Park. El formato es similar al del problema de la ruta más corta que se presentó en la figura 9.4. Los arcos se enumeran en las columnas B y C, y las capacidades de arco correspondientes se proporcionan en la columna F. Como las variables de decisión son los flujos a través de los arcos respectivos, estas cantidades se introducen en las celdas cambiantes Flujo (D4:D15). Al emplear las ecuaciones dadas en la esquina inferior derecha de la figura, estos flujos se usan para calcular el flujo neto generado en cada nodo (vea las columnas H e I). En los nodos de trasbordo (A, B, C, D y E) se requiere que estos flujos netos sean 0, como se indica en el primer conjunto de restricciones (I5:I9 = DemandaDeSuministros) en el cuadro de diálogo de Solver. El segundo conjunto de restricciones (Flujo ≤ Capacidad) especifica las restricciones de capacidad de arco.

■ **FIGURA 9.10**
Corte mínimo del problema de flujo máximo de Seervada Park.



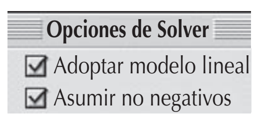


	A	B	C	D	E	F	G	H	I	J	K
1	Problema de flujo máximo de Seervada Park										
2											Suministro/
3		De	A	Flujo		Capacidad		Nodos	Flujo neto		Demanda
4		O	A	4		5		O	14		
5		O	B	7		7		A	0	=	0
6		O	C	3		4		B	0	=	0
7		A	B	1		1		C	0	=	0
8		A	D	3		3		D	0	=	0
9		B	C	0		2		E	0	=	0
10		B	D	4		4		T	-14		
11		B	E	4		5					
12		C	E	3		4					
13		D	T	8		9					
14		E	D	1		1					
15		E	T	6		6					
16											
17			Flujo máximo	14							



	I
3	Flujo neto
4	=SUMASI(De,H4,Flujo)-SUMASI(A,H4,Flujo)
5	=SUMASI(De,H5,Flujo)-SUMASI(A,H5,Flujo)
6	=SUMASI(De,H6,Flujo)-SUMASI(A,H6,Flujo)
7	=SUMASI(De,H7,Flujo)-SUMASI(A,H7,Flujo)
8	=SUMASI(De,H8,Flujo)-SUMASI(A,H8,Flujo)
9	=SUMASI(De,H9,Flujo)-SUMASI(A,H9,Flujo)
10	=SUMASI(De,H10,Flujo)-SUMASI(A,H10,Flujo)

Nombre de rango	Celdas
Capacidad	F4:F15
Flujo	D4:D15
De	B4:B15
FlujoMax	D17
FlujoNeto	I4:I10
Nodos	H4:H1 0
SuministroDemanda	K5:K9
A	C4:C15



	C	D
17	Flujo máximo	=I4

■ **FIGURA 9.11**
Formulación en hoja de cálculo del problema de flujo máximo de Seervada Park, en la cual las celdas que cambian Flujo (D4:D15) muestran la solución óptima que se obtiene mediante Excel Solver, mientras que la celda objetivo FlujoMax (D17) proporciona el flujo máximo que resulta a través de la red. La red que se encuentra a continuación de la hoja de cálculo muestra el problema de flujo máximo original que se presentó antes en la figura 9.6.

La cantidad total de flujo de la fuente (nodo O) al destino (nodo T) es igual al flujo generado en la fuente (celda 14), de manera que la celda objetivo FlujoMax(D17) se establece igual a 14. Después de especificar *maximización* de la celda objetivo en Solver y de hacer clic en resolver, se obtiene la solución óptima que se presenta en Flujo (D4:D15).

9.6 PROBLEMA DEL FLUJO DE COSTO MÍNIMO

El problema del flujo de costo mínimo tiene una posición central entre los modelos de optimización de redes; primero, abarca una clase amplia de aplicaciones y, segundo, su solución es muy eficiente. Igual que el problema del flujo máximo, toma en cuenta un flujo en una red con capacidades de arco limitadas. Igual que el problema de la ruta más corta, considera un costo (o distancia) del flujo a través de un arco. Igual que el problema de transporte o el de asignación del capítulo 8, puede manejar varios orígenes (nodos fuente) y varios destinos (nodos demanda) del flujo, de nuevo con costos asociados. En realidad, estos cuatro problemas son casos especiales del problema del flujo de costo mínimo, como se demostrará en esta sección.

La razón por la que el problema del flujo de costo mínimo se puede resolver de modo tan eficiente es que se puede formular como un problema de programación lineal y es posible resolverlo con una versión simplificada del método símplex llamada *método símplex de redes*. En la siguiente sección se describirá este algoritmo.

A continuación se describe el problema del flujo de costo mínimo.

1. La red es una red *dirigida* y *conexa*.
2. *Al menos uno* de los nodos es un *nodo fuente*.
3. *Al menos uno* de los nodos es un *nodo demanda*.
4. El resto de los nodos son *nodos de trasbordo*.
5. Se permite el flujo a través de un arco sólo en la dirección que indica la flecha, donde la cantidad máxima de flujo está dada por la capacidad del arco. (Si el flujo puede ocurrir en ambas direcciones, debe representarse por un par de arcos con direcciones opuestas.)
6. La red tiene suficientes arcos con suficiente capacidad para permitir que todos los flujos generados por los *nodos fuente* lleguen a los *nodos demanda*.
7. El costo del flujo a través del arco es *proporcional* a la cantidad de ese flujo, donde se conoce el costo por unidad.
8. El objetivo es minimizar el costo total de enviar el suministro disponible a través de la red para satisfacer la demanda dada. (Un objetivo alternativo es maximizar la ganancia total del envío.)

Algunas aplicaciones

Tal vez el tipo más importante de aplicación del problema del flujo de costo mínimo es en la operación de la red de distribución de una compañía. Como se resume en el primer renglón de la tabla 9.3, este tipo de aplicación siempre incluye determinar un plan para enviar bienes desde las *fuentes* (fábricas, etc.) a las *instalaciones de almacenamiento intermedias* (según se necesite) y después a los *clientes*.

En el caso de algunas aplicaciones de los problemas del flujo de costo mínimo, todos los nodos de trasbordo son *instalaciones de procesamiento* y no almacenes. Éste es el caso de la *administración de desechos sólidos*, indicado en el segundo renglón de la tabla 9.3. En ese problema, el flujo de materiales a través de la red comienza en las fuentes de desechos sólidos, luego va a las instalaciones para procesar estos materiales de desecho y convertirlos en una forma adecuada para el relleno y después se envía a los diferentes rellenos. Sin embargo, el objetivo todavía es determinar el plan de flujo que minimice el costo total, donde el costo ahora se refiere al embarque y al procesamiento.

En otras aplicaciones, los *nodos demanda* pueden ser instalaciones de procesamiento. Por ejemplo, en el tercer renglón de la tabla 9.3, el objetivo es encontrar el plan de costo mínimo para obtener suministros de varios proveedores posibles, almacenar los bienes —según se requiera—, y luego enviar los suministros a las instalaciones de procesamiento de la compañía (fábricas, etc.). Como la cantidad total que pueden suministrar los proveedores es mayor que las necesidades de la compañía, la red incluye un *nodo de demanda ficticio* que recibe (sin costo) toda la capacidad no usada de los proveedores.

El siguiente tipo de aplicación de la tabla 9.3 —coordinación de mezclas de productos en las plantas— ilustra que los arcos pueden representar algo distinto a un canal de envío de un flujo fí-

TABLA 9.3 Aplicaciones comunes del problema del flujo de costo mínimo

Tipo de aplicación	Nodos fuente	Nodos de trasbordo	Nodos demanda
Operación de una red de distribución	Fuentes de bienes	Almacenes intermedios	Clientes
Administración de desechos sólidos	Fuentes de desechos sólidos	Instalaciones de procesamiento	Rellenos
Operación de una red de suministros	Agentes de ventas	Almacenes intermedios	Instalaciones de procesamiento
Coordinación de mezclas de productos en plantas	Plantas	Producción de un artículo específico	Mercado del producto específico
Administración de flujo de efectivo	Fuentes de efectivo en tiempos específicos	Opciones de inversión a corto plazo	Necesidad de efectivo en tiempos específicos

Recuadro de aplicación

Un problema que representa un verdadero reto en particular para cualquier compañía aérea diariamente consiste en cómo compensar de manera efectiva las demoras en los horarios de sus vuelos. El mal tiempo, así como los problemas mecánicos, pueden retrasar las salidas y llegadas de los vuelos. Cada retraso o cancelación que involucra a una aeronave en particular puede provocar demoras y cancelaciones subsecuentes debido a que la aeronave no está disponible a la hora correcta para sus siguientes vuelos programados.

Dichos retrasos y cancelaciones pueden causar que se reasigne la tripulación a los vuelos y que se reajusten los planes para los que se usarán las aeronaves a fin de realizar sus respectivos vuelos. El recuadro de aplicación de la sección 2.2 describe cómo Continental Airlines lideró el camino en la aplicación de la investigación de operaciones al problema de reasignar tripulaciones a los vuelos de manera rápida y eficiente. Sin embargo, es necesario utilizar otro método para resolver el problema de la reasignación rápida de aeronaves a los vuelos.

Una aerolínea cuenta con dos formas principales para reasignar aeronaves a los vuelos para compensar los retrasos o cancelaciones. Una de ellas consiste en intercambiar aeronaves de tal forma que un avión que estuviera programado para un vuelo posterior pueda hacerse cargo del vuelo retrasado o cancelado. La otra consiste en usar un avión extra (a menudo después de que éste ha volado) para reemplazar la aeronave retrasada o cancelada. Sin embargo, representa un verdadero reto llevar a cabo las decisiones correctas de manera rápida cuando se presentan infinidad de retrasos y cancelaciones en un día.

United Airlines ha sido líder en la forma de aplicar la investigación de operaciones a este problema. Esto lo lleva a cabo mediante la formulación y resolución del problema como un *problema de flujo de costo mínimo*, donde cada nodo en la red representa un aeropuerto y cada arco la ruta del vuelo. El objetivo del modelo es mantener a las aeronaves volando a través de la red de manera tal que se minimice el costo asociado con los retrasos y las cancelaciones. En el momento en que el subsistema de supervisión de estado envía una señal de alerta al controlador de operaciones de retrasos y cancelaciones inminentes, el controlador genera los datos de entrada necesarios al modelo y éste resuelve el problema a fin de obtener un plan de operaciones actualizado en cuestión de minutos. Esta aplicación de problema del flujo de costo mínimo ha representado *una reducción en retrasos de pasajeros de alrededor de 50 por ciento*.

Fuente: A. Rakshit, N. Krishnamurthy y G. Yu: "System Operations Advisor: A Real-Time Decision Support System for Managing Airline Operations at United Airlines", *Interfaces*, **26**(2): 50-58, marzo-abril, 1996. (En el sitio de internet de este libro [www.mhhe.com/hillier] se proporciona una liga hacia este artículo.)

sico de materiales. Esta aplicación se refiere a una compañía con varias plantas (nodos fuente) que pueden producir los mismos productos pero con diferentes costos. Cada arco que sale de un nodo fuente representa la producción de uno de los productos posibles de esa planta, donde el arco conduce a un nodo de trasbordo que corresponde al producto. En consecuencia, este nodo de trasbordo tiene un arco que llega desde cada planta capaz de producir este producto, y después los arcos que salen de este nodo van a los respectivos clientes (nodos de demanda). El objetivo es determinar cómo dividir la capacidad de producción de cada planta entre los productos para minimizar el costo total de satisfacer las demandas de los distintos productos.

La última aplicación de la tabla 9.3 (administración de flujo de efectivo), ejemplifica que diferentes nodos pueden representar un evento que ocurre en distintos momentos. En este caso, cada nodo fuente representa un tiempo específico (periodo) en que la compañía dispone de dinero (por madurez de inversiones, cuentas por cobrar, venta de acciones, préstamos, etc.). El suministro de cada nodo es la cantidad de efectivo que estará disponible. De manera similar, cada nodo destino representa un tiempo específico (periodo) en que la compañía deberá usar su reserva de efectivo. La demanda de cada uno de estos nodos es la cantidad de efectivo que necesitará. El objetivo es maximizar el ingreso de la compañía debido a estas inversiones entre los tiempos en que se dispone de efectivo y se usa. En consecuencia, los nodos de trasbordo representan la opción de una inversión a corto plazo —es decir, la compra de certificados de depósito en un banco— por un periodo específico. La red resultante tendrá una serie de flujos que representan un programa de efectivo disponible, invertido y utilizado cuando la inversión madura.

Formulación del modelo

Considere una red conexa dirigida en la que los n nodos incluyen al menos un nodo origen y un nodo destino. Las variables de decisión son

$$x_{ij} = \text{flujo a través del arco } i \rightarrow j,$$

y la información dada incluye

c_{ij} = costo por unidad de flujo a través del arco $i \to j$,
u_{ij} = capacidad del arco $i \to j$,
b_i = flujo neto generado por el nodo i.

El valor de b_i depende de la naturaleza del nodo i, donde

$b_i > 0$ si i es un nodo fuente,
$b_i < 0$ si i es un nodo demanda,
$b_i = 0$ si i es un nodo de trasbordo.

El objetivo es minimizar el costo total de enviar los recursos disponibles a través de la red para satisfacer la demanda.

Si se usa la convención de que las sumas se toman sólo sobre arcos existentes, la formulación de programación lineal de este problema es

$$\text{Minimizar} \qquad Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}x_{ij},$$

sujeta a

$$\sum_{j=1}^{n} x_{ij} - \sum_{j=1}^{n} x_{ji} = b_i, \qquad \text{para cada nodo } i,$$

y

$$0 \le x_{ij} \le u_{ij}, \qquad \text{para cada arco } i \to j.$$

La primera suma de las *restricciones de los nodos* representa el flujo total que *sale* del nodo i mientras que la segunda representa el flujo total que *entra* al nodo i; por tanto, la diferencia es el flujo neto generado en este nodo.

El patrón de los coeficientes de estas restricciones de nodo es una característica importante de los problemas de flujo de costo mínimo. No siempre es fácil reconocer un problema de flujo de costo mínimo, pero al formular (o reformular) un problema de manera que sus coeficientes de restricción tengan este patrón es una buena forma de hacerlo. Lo anterior permite resolver el problema de manera muy eficiente mediante el método símplex de redes.

En algunas aplicaciones es necesario tener una cota inferior $L_{ij} > 0$ para el flujo que pasa por cada arco $i \to j$. Cuando esto ocurre se hace una conversión de variables, $x'_{ij} = x_{ij} - L_{ij}$, donde x_{ij} se sustituye por $x'_{ij} + L_{ij}$ en todo el modelo, a fin de ajustarlo al formato anterior con restricciones de no negatividad.

No se garantiza que el problema tenga *soluciones factibles*, pues esto depende en parte de qué arcos están presentes en la red y de sus capacidades. De cualquier manera, para una red diseñada en forma razonable, la condición necesaria más importante es la siguiente.

Propiedad de soluciones factibles: una condición necesaria para que un problema de flujo de costo mínimo tenga soluciones factibles es que

$$\sum_{i=1}^{n} b_i = 0.$$

Es decir, que el flujo total generado por los nodos origen es igual al flujo total absorbido por los nodos destino.

Si los valores de b_i, que se dan en alguna aplicación violan esta condición, la interpretación más común es que los recursos o las demandas —lo que se tenga en exceso— representan en realidad cotas superiores y no cantidades exactas. Cuando esta situación surgió en el problema de transporte de la sección 8.1, se agregó un destino ficticio para recibir los recursos que sobraban o bien se añadió un origen ficticio para enviar el exceso de demanda. El paso análogo es agregar un nodo de demanda ficticio para absorber el exceso de recursos —se agregan arcos con $c_{ij} = 0$ desde todos

los nodos origen hasta este nodo—, o bien agregar un nodo origen ficticio para generar un flujo equivalente al exceso de demanda —se agregan arcos con $c_{ij} = 0$ de este nodo a todos los nodos de demanda—.

En muchas aplicaciones, las cantidades b_i y u_{ij} tendrán valores *enteros* y la solución requerirá que las cantidades de flujo x_{ij} también sean enteras. Por fortuna, igual que para el problema de transporte, este tipo de solución está garantizada sin tener que establecer restricciones enteras de manera explícita sobre las variables. Esto se debe a la siguiente propiedad.

> **Propiedad de soluciones enteras:** en el caso de los problemas de flujo de costo mínimo en donde toda b_i y u_{ij} tienen un valor entero, todas las variables de *cada* solución básica factible (BF), incluida la óptima, tendrán también valores enteros.

Un ejemplo

En la figura 9.12 se muestra un ejemplo del problema del flujo de costo mínimo. Esta red, en realidad, es la *red de distribución* del problema de Distribution Unlimited Co., que se presentó en la sección 3.4 (vea la figura 3.13). Las cantidades que se presentaron en la figura 3.13 proporcionan los valores de b_i, c_{ij} y u_{ij} que se muestran aquí. Los valores de b_i de la figura 9.12 están entre corchetes cerca de los nodos; entonces, los nodos origen ($b_i > 0$) son A y B (las dos fábricas de la compañía), los nodos destino ($b_i < 0$) son D y E (los dos almacenes), y el único nodo de trasbordo ($b_i = 0$) es C (un centro de distribución). Los valores c_{ij} se muestran junto a los arcos. En este ejemplo, todos menos dos de los arcos tienen capacidades que exceden el flujo total generado (90), de manera que $u_{ij} = \infty$ para cualquier propósito práctico. Las dos excepciones son el arco $A \to B$, donde $u_{AB} = 10$ y el arco $C \to E$ que tiene $u_{CE} = 80$.

El modelo de programación lineal de este ejemplo es

$$\text{Minimizar} \quad Z = 2x_{AB} + 4x_{AC} + 9x_{AD} + 3x_{BC} + x_{CE} + 3x_{DE} + 2x_{ED},$$

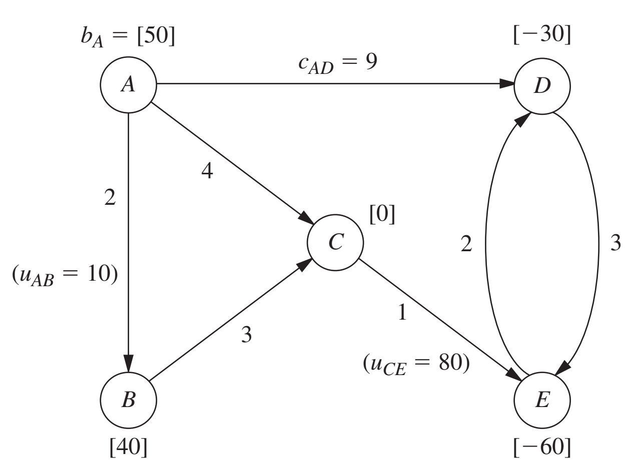
sujeta a

$$\begin{aligned}
x_{AB} + x_{AC} + x_{AD} \qquad\qquad\qquad\qquad\qquad &= 50\\
-x_{AB} \qquad\quad + x_{BC} \qquad\qquad\qquad\qquad &= 40\\
- x_{AC} \qquad - x_{BC} + x_{CE} \qquad\qquad\quad &= 0\\
- x_{AD} \qquad\qquad + x_{DE} - x_{ED} &= -30\\
- x_{CE} - x_{DE} + x_{ED} &= -60
\end{aligned}$$

y

$$x_{AB} \le 10, \quad x_{CE} \le 80, \quad \text{toda } x_{ij} \ge 0.$$

Ahora observe el patrón de coeficientes de cada variable en el conjunto de cinco *restricciones* de nodo (restricciones de igualdad). Cada variable tiene exactamente *dos* coeficientes distintos de



■ **FIGURA 9.12**
El problema de la Distribution Unlimited Co., formulado como uno de flujo de costo mínimo.

cero, uno es +1 y el otro −1. Este patrón aparece en *todos* los problemas de flujo de costo mínimo y es esta estructura especial la que lleva a la propiedad de soluciones enteras.

Otra consecuencia de esta estructura especial es que una (cualquiera) de las restricciones de nodo es *redundante*. La razón es que si se suman todas estas ecuaciones sólo se obtienen ceros en ambos lados —si se supone que existen soluciones factibles para que las b_i sumen cero—, por lo que el negativo de cualquier ecuación es igual a la suma de las demás. Con $n - 1$ restricciones de arco no redundantes, estas ecuaciones proporcionan exactamente $n - 1$ variables básicas para una solución básica factible. En la siguiente sección se verá que el método símplex de redes trata las restricciones $x_{ij} \leq u_{ij}$ como simétricas de las restricciones de no negatividad; por tanto, el número *total* de variables básicas es $n - 1$. Esto conduce a una correspondencia directa entre los $n - 1$ arcos de un *árbol de expansión* y las $n - 1$ variables básicas. Se hablará más sobre este tema más adelante.

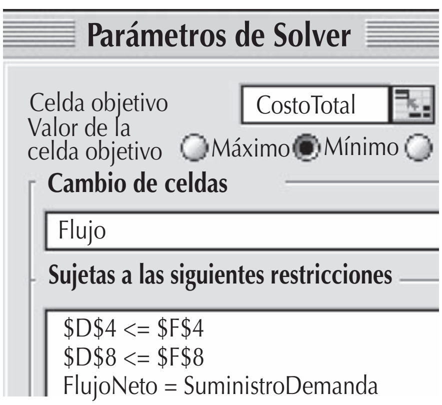
Uso de Excel para formular y resolver problemas de flujo de costo mínimo

Excel proporciona una manera conveniente de formular y resolver problemas de flujo de costo mínimo como éste, y algunos más grandes. La figura 9.13 muestra cómo se pueden alcanzar las

■ **FIGURA 9.13**

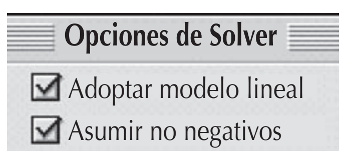
Formulación en hoja de cálculo del problema del flujo de costo mínimo de Distribution Unlimited Co., en la cual las celdas que cambian Embarque (D4:D10) muestran la solución óptima que se obtiene mediante Excel Solver, mientras que la celda objetivo CostoTotal (D12) proporciona el costo total del flujo de envíos a través de la red.

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Problema de flujo de costo mínimo de Distribution Unlimited Co.**											
2												**Suministro/**
3		**De**	**A**	**Embarque**		**Capacidad**	**Costo unit.**		**Nodos**	**Flujo neto**		**Demanda**
4		A	B	0		10	2		A	50	=	50
5		A	C	40			4		B	40	=	40
6		A	D	10			9		C	0	=	0
7		B	C	40			3		D	-30	=	−30
8		C	E	80	8	0	1		E	-60	=	−60
9		D	E	0			3					
10		E	D	20			2					
11												
12			**Costo total**	490								



Nombre de rango	Celdas
Capacidad	F4:F10
De	B4:B10
FlujoNeto	J4:J8
Nodos	I4:I8
Embarque	D4:D10
SuministroDemanda	L4:L8
A	C4:C10
CostoTotal	D12
CostoUnitario	G4:G10

	J
3	**Flujo neto**
4	=SUMASI(De,I4,Embarque)-SUMASI(A,I4,Embarque)
5	=SUMASI(De,I5,Embarque)-SUMASI(A,I5,Embarque)
6	=SUMASI(De,I6,Embarque)-SUMASI(A,I6,Embarque)
7	=SUMASI(De,I7,Embarque)-SUMASI(A,I7,Embarque)
8	=SUMASI(De,I8,Embarque)-SUMASI(A,I8,Embarque)



	C	D
12	**Costo total**	=SUMAPRODUCTO(D4:D10,G4:G10)

soluciones. El formato es casi el mismo que el de la figura 9.11 en el caso de un problema de flujo máximo. Una de las diferencias es que ahora deben incluirse los costos unitarios (c_{ij}), en la columna *G*. Como se especifican los valores de b_i de cada nodo, se requieren las restricciones de flujo neto para todos los nodos. Sin embargo, en realidad sólo dos arcos necesitan restricciones de capacidad. La celda objetivo CostoTotal(D12) ahora proporciona el costo total del flujo (embarques) a través de la red (vea su ecuación en la parte inferior de la figura), de manera que el objetivo especificado en el cuadro de diálogo de Solver es *minimizar* esta cantidad. En esta hoja de cálculo, las celdas cambiantes Embarque (D4:D10) muestran la solución óptima que se obtiene después de hacer clic en resolver.

En el caso de problemas más grandes de flujo de costo mínimo, el *método símplex de redes* descrito en la siguiente sección proporciona un procedimiento de solución mucho más eficiente. También es una opción atractiva para resolver varios casos especiales del problema del flujo de costo mínimo descrito en seguida. Es común que se incluya este algoritmo en los paquetes de programación matemática. Por ejemplo, es una de las opciones de CPLEX.

Pronto se resolverá este ejemplo con el método símplex de transporte. Pero primero se analizará la manera en que algunos casos especiales se ajustan al formato del problema del flujo de costo mínimo.

Casos especiales

El problema de transporte. Para formular el problema de transporte presentado en la sección 8.1 como un problema de flujo de costo mínimo, se proporciona un *nodo de recursos* para cada *origen* y un *nodo de demanda* para cada *destino*, pero no se incluyen nodos de trasbordo en la red. Todos los arcos son dirigidos desde el nodo de recursos hasta el nodo de demanda, en donde distribuir x_{ij} unidades del origen i al destino j corresponde a un flujo de x_{ij} a través del arco $i \rightarrow j$. El costo c_{ij} por unidad distribuida se convierte en el costo c_{ij} por unidad de flujo. Como el problema de transporte no impone restricciones de cota superior sobre las x_{ij} individuales, todas las $u_{ij} = \infty$.

Al utilizar esta formulación para el problema de transporte de la P & T Co., que se presentó en la tabla 8.2 se llega a la red que se muestra en la figura 8.2. La red correspondiente para el problema general de transporte se muestra en la figura 8.3.

El problema de asignación. Debido a que el problema de asignación que se presentó en la sección 8.3 es un tipo especial de problema de transporte, su formulación como un problema del flujo de costo mínimo se ajusta al mismo formato. Los factores adicionales son que 1) el número de nodos de recursos es igual al número de nodos de demanda, 2) $b_i = 1$ en cada nodo de recursos y 3) $b_i = -1$ en cada nodo de demanda.

En la figura 8.5 se muestra esta formulación del problema general de asignación.

El problema de trasbordo. En realidad, este caso especial incluye todas las características generales del problema del flujo de costo mínimo excepto que no tiene capacidades (finitas) en los arcos. En consecuencia, cualquier flujo de costo mínimo en el que cada arco puede conducir cualquier cantidad de flujo deseada se llama también problema de trasbordo.

Por ejemplo, el problema de la Distribution Unlimited Co., que se presentó en la figura 9.13 sería un problema de trasbordo si se eliminaran las cotas superiores sobre los flujos a través de los arcos $A \rightarrow B$ y $C \rightarrow E$.

Con frecuencia, los problemas de trasbordo surgen como generalizaciones de problemas de transporte en los cuales las unidades distribuidas desde cada origen a cada destino pueden pasar primero por puntos intermedios. Estos puntos intermedios pueden incluir otros orígenes y destinos, al igual que puntos de transferencia adicionales que se deben representar por nodos de trasbordo en la red del problema. Por ejemplo, el problema de la Distribution Unlimited Co., se puede considerar como una generalización de un problema de transporte con dos orígenes —las dos fábricas representadas por los nodos *A* y *B* en la figura 9.13—, dos destinos —los dos almacenes simbolizados por los nodos *D* y *E*— y un punto de transferencia intermedio adicional —el centro de distribución que representa el nodo *C*—.

(En el capítulo 23 del sitio en internet de este libro se incluye una presentación más profunda del problema de trasbordo.)

El problema de la ruta más corta. A continuación se considerará la versión del problema de la ruta más corta que se presentó en la sección 9.3 (para encontrar la ruta más corta desde un origen a un destino a través de una red *no dirigida*). Para formular este problema como un problema del flujo de costo mínimo se establece el origen como un nodo de recursos con una cantidad de 1, el destino como un nodo de demanda con una demanda de 1 y el resto de los nodos como nodos de trasbordo. Debido a que la red del problema de la ruta más corta es no dirigida, mientras que se supone que el problema del flujo de costo mínimo tiene una red dirigida, cada ligadura se sustituye por un par de arcos con direcciones opuestas (dibujados como una sola línea con cabezas de flechas en ambas terminales). Las únicas excepciones son que no hay necesidad de preocuparse por arcos *que llegan* al nodo origen ni que *salen* del nodo destino. La distancia entre los nodos i y j se convierte en unidades de costo c_{ij} o c_{ji} para el flujo en cualquier dirección entre estos nodos. Igual que en los casos especiales anteriores, no se imponen capacidades a los arcos por lo que toda $u_{ij} = \infty$.

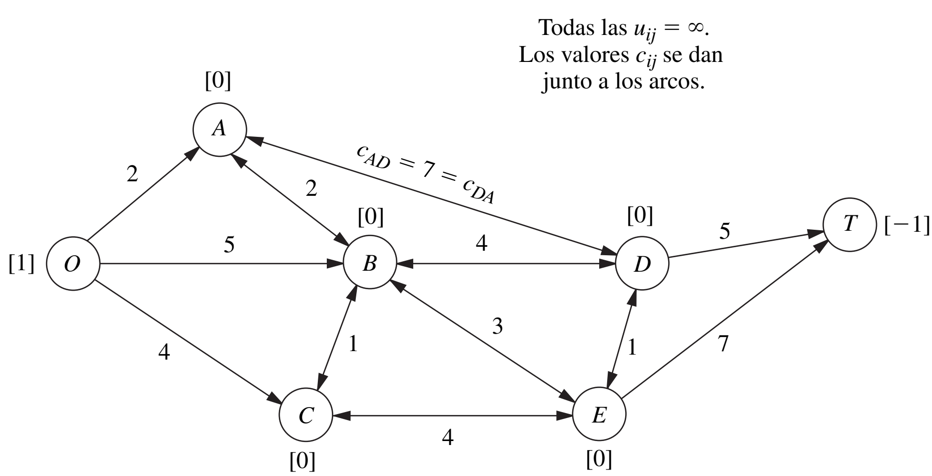
En la figura 9.14 se describe esta formulación para el problema de la ruta más corta de Seervada Park que se presentó en la figura 9.1, en la cual ahora los números cerca de las líneas representan el costo unitario del flujo en cualquier dirección.

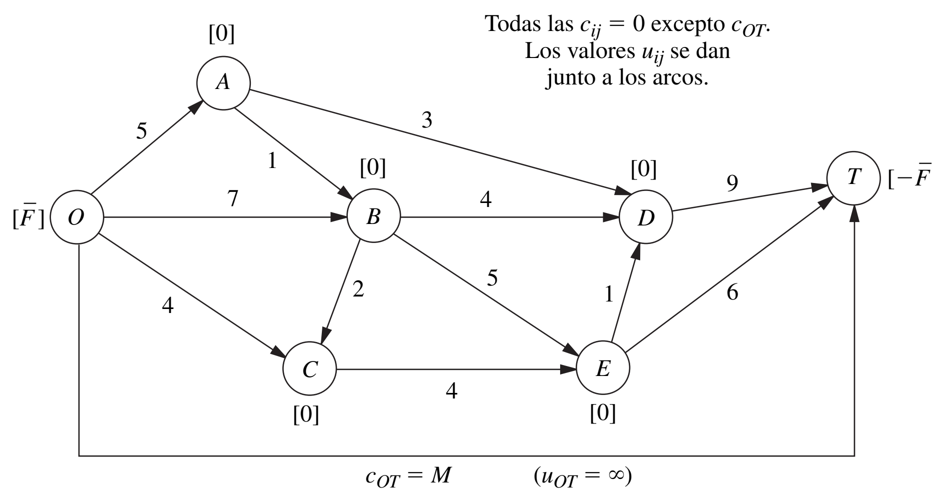
Problema de flujo máximo. El último caso especial que se considerará es el problema del flujo máximo descrito en la sección 9.5. En este caso, la red ya tiene un nodo origen (fuente), un nodo de demanda (destino) y varios nodos de trasbordo, al igual que varios arcos y capacidades en los arcos. Sólo se necesitan tres ajustes para que este problema quede en el formato del problema del flujo de costo mínimo. Uno es hacer $c_{ij} = 0$ para todos los arcos existentes de manera que se refleje la ausencia de costos en el problema del flujo máximo. Otro es elegir una cantidad $\overline{F}$ que sea una cota superior segura sobre el flujo factible máximo a través de la red y después asignar una cantidad de recursos y de demanda $\overline{F}$ a los nodos de recursos y de demanda, respectivamente. (Como todos los otros nodos son de trasbordo, de manera automática tienen $b_i = 0$.) El tercer ajuste es agregar un arco que va en forma directa desde el nodo de recursos hasta el nodo de demanda y asignarle un costo unitario arbitrariamente grande $c_{ij} = M$ y una capacidad de arco ilimitada ($u_{ij} = \infty$). Debido a este costo unitario positivo de este arco y al costo unitario de cero de los *otros*, el problema del flujo de costo mínimo mandará el máximo flujo factible a través de los *otros* arcos, con lo que se logra el objetivo del problema del flujo máximo.

La aplicación de esta formulación al problema del flujo máximo en Seervada Park que se muestra en la figura 9.6, conduce a la red dada en la figura 9.15, en la cual los números ubicados cerca de los arcos originales son sus capacidades.

Comentarios finales. Con excepción del problema de trasbordo, cada uno de estos casos especiales ha sido el centro de atención de una sección anterior, ya sea en este capítulo o en el 8. Cuando se presentó cada uno por primera vez, se habló de algoritmos especializados para resolverlos de manera muy eficiente. En consecuencia, no es necesario reformular estos casos especiales para

■ **FIGURA 9.14**
Formulación del problema de la ruta más corta de Seervada Park como un problema de flujo de costo mínimo.





■ **FIGURA 9.15**
Formulación del problema del flujo máximo de Seervada Park como un problema de flujo de costo mínimo.

que se ajusten al problema del flujo de costo mínimo a fin de resolverlos. Sin embargo, cuando no se dispone de un paquete de computadora para el algoritmo especializado, resulta razonable utilizar el método símplex de redes. En realidad, algunas aplicaciones recientes de este algoritmo han sido tan poderosas que en la actualidad proporcionan una alternativa excelente en lugar de los algoritmos especializados.

El hecho de que estos cinco problemas sean casos especiales del problema del flujo de costo mínimo también es de interés por otras razones. Una es que la teoría que subyace al problema del flujo de costo mínimo y al método símplex de redes proporciona un enfoque unificador para todos estos casos especiales. Otra razón es que algunas aplicaciones del problema del flujo de costo mínimo incluyen características de uno o más casos especiales, por lo que es importante saber cómo reformular estas características dentro del contexto más amplio del problema general.

9.7 MÉTODO SÍMPLEX DE REDES

El método símplex de redes es una versión muy simplificada del método símplex para resolver problemas de flujo de costo mínimo. Como tal, realiza los mismos pasos básicos en cada iteración —encontrar la variable básica entrante, determinar la variable básica que sale y obtener la nueva solución BF— con el fin de mover la solución BF actual a una adyacente mejor. Sin embargo, ejecuta estos pasos en una forma que explota la estructura especial de la red del problema sin necesidad alguna de la tabla símplex.

Es posible observar algunas similitudes entre el método símplex de redes y el método símplex de transporte que se presentó en la sección 8.2. En realidad, ambos son versiones simplificadas del método símplex que proporcionan algoritmos alternativos para resolver problemas de transporte de manera parecida. El método símplex de redes extiende estas ideas para resolver también otros tipos de problemas de flujo de costo mínimo.

En esta sección se presenta una descripción abreviada del método símplex de redes que centra la atención exactamente en los conceptos principales. Se omiten ciertos detalles necesarios para manipularlo en computadora, entre otros, cómo construir una solución inicial BF y cómo realizar ciertos cálculos —encontrar la variable básica entrante— de una manera eficiente. Estos detalles se proporcionan en otros libros más especializados, como las referencias 1 y 3.

Incorporación de la técnica de la cota superior

El primer concepto es incorporar la técnica de la cota superior que se presentó en la sección 7.3 para manejar con eficiencia las restricciones de capacidad de arco $x_{ij} \leq u_{ij}$. De esta forma, en lugar de tratar estas restricciones como funcionales, se manejan como restricciones de *no negatividad*. Por lo tanto, sólo se toman en cuenta para determinar la variable básica que sale. En particular, a

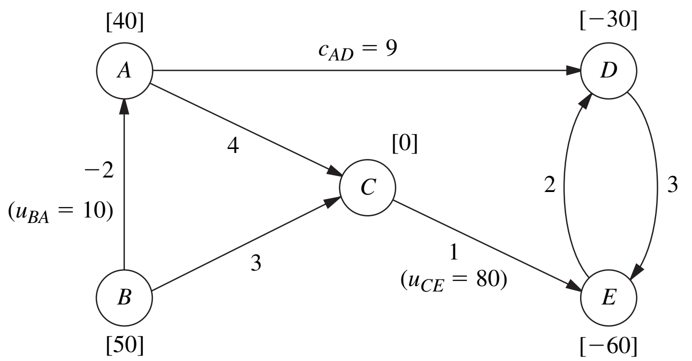
medida que se hace crecer la variable básica entrante desde cero, la variable básica que sale es la *primera* variable básica que llega, ya sea a su cota inferior (0) o a su cota superior (u_{ij}). Una variable no básica en su cota superior $x_{ij} = u_{ij}$, se reemplaza por $x_{ij} = u_{ij} - y_{ij}$ de manera que $y_{ij} = 0$ se convierte en la variable no básica. Vea más detalles en la sección 7.3

En el contexto actual, y_{ij} tiene una interpretación interesante en la red. Siempre que y_{ij} se convierta en una variable básica con valor estrictamente positivo ($\leq u_{ij}$), se puede pensar en este valor como en un flujo del nodo *j* al nodo *i* —o sea, en la dirección "equivocada" a través del arco $i \rightarrow j$— que en realidad *cancela* la cantidad de flujo previamente asignado ($x_{ij} = u_{ij}$) del nodo *i* al nodo *j*. Así, cuando $x_{ij} = u_{ij}$ se reemplaza con $x_{ij} = u_{ij} - y_{ij}$, también se reemplaza el arco real $i \rightarrow j$ por su **arco inverso** $j \rightarrow i$, en donde este nuevo arco tiene capacidad u_{ij} —la cantidad máxima del flujo $x_{ij} = u_{ij}$ que se puede cancelar— y costo unitario $-c_{ij}$ —puesto que cada unidad de flujo cancelada ahorra c_{ij}—. Para reflejar el flujo de $x_{ij} = u_{ij}$ a través del arco eliminado, se cambia esta cantidad de flujo neto generada del nodo *i* al nodo *j* al *disminuir* b_i en u_{ij} unidades e *incrementar* b_j en u_{ij} unidades. Después, si la y_{ij} se convierte en la variable básica que sale al alcanzar su cota superior, $y_{ij} = u_{ij}$ se reemplaza por $y_{ij} = u_{ij} - x_{ij}$, con $x_{ij} = 0$ como la nueva variable no básica, con lo que el proceso anterior se invierte —el arco $j \rightarrow i$ se sustituye por el arco $i \rightarrow j$, etc.— para regresar a la configuración original.

Para ilustrar este proceso, considere el problema del flujo de costo mínimo que se muestra en la figura 9.12. Mientras que el método símplex de redes genera una sucesión de soluciones BF, suponga que en alguna iteración, x_{AB} se convirtió en la variable básica que sale al alcanzar su cota superior de 10. En consecuencia, $x_{AB} = 10$ se sustituye por $x_{AB} = 10 - y_{AB}$, de manera que $y_{AB} = 0$ se convierte en la nueva variable no básica. Al mismo tiempo, el arco $B \rightarrow A$ sustituye al arco $A \rightarrow B$ (con y_{AB} como su flujo), y se asigna a este nuevo arco una capacidad de 10 y un costo unitario de -2. Para tomar en cuenta que $x_{AB} = 10$, se disminuye b_A de 50 a 40 y se aumenta b_B de 40 a 50. En la figura 9.16 se muestran los ajustes que resultan.

Con este mismo ejemplo, se ilustrará todo el método símplex de redes a partir de $y_{AB} = 0$ ($x_{AB} = 10$) como variable no básica y mediante el uso de la figura 9.16. Una iteración posterior mostrará que x_{CE} alcanza su cota superior de 80 y se sustituye por $x_{CE} = 80 - y_{CE}$, etc., y entonces en la siguiente iteración ocurre que y_{AB} alcanza su cota superior de 10. Se podrá observar que estas operaciones se realizan de manera directa sobre la red, así que no será necesario utilizar las etiquetas x_{ij} o y_{ij} para designar los flujos en los arcos y ni siquiera se tendrá que saber qué arcos son arcos *reales* y cuáles son *inversos* (excepto cuando se registre la solución final). El uso de esta técnica de la cota superior deja las *restricciones de los nodos* —el flujo que sale menos el flujo que entra = b_i— como las únicas restricciones funcionales. Los problemas de flujo de costo mínimo tienden a tener muchos más arcos que nodos, por lo que el número de restricciones funcionales que resulta es, por lo general, una pequeña fracción de lo que sería si se incluyeran las restricciones de las capacidades de los arcos. El tiempo de cálculo que emplea el método símplex crece relativamente rápido a medida que crece el número de restricciones funcionales, pero aumenta despacio con el número de variables (o el número de restricciones de acotamiento sobre estas variables). Por lo tanto, la incorporación de la técnica de la cota superior tiende a proporcionar ahorros considerables en tiempo de cálculo.

FIGURA 9.16
Red ajustada para el ejemplo cuando la técnica de la cota superior conduce a reemplazar $x_{AB} = 10$ por $x_{AB} = 10 - y_{AB}$.



Sin embargo, no se necesita esta técnica para problemas de flujo de costo mínimo con arcos *no capacitados* —tema que incluye todos los casos especiales, menos el último, considerados en la sección anterior—, en donde no existen las restricciones de capacidad de arco.

Correspondencia entre las soluciones BF y los árboles de expansión factibles

El concepto más importante que apoya el método símplex de redes es su representación como una red de las *soluciones BF*. Recuerde que en la sección 9.6 se estableció que con n nodos, toda solución básica factible tiene $(n - 1)$ variables básicas, y que cada variable básica x_{ij} representa el flujo por el arco $i \rightarrow j$. Estos $(n - 1)$ arcos se conocen como **arcos básicos**. (De manera similar, los arcos que corresponden a variables *no básicas*, $x_{ij} = 0$ o $y_{ij} = 0$, se llaman **arcos no básicos**.)

Una propiedad importante de los arcos básicos es que nunca forman *ciclos* no dirigidos. (Esta propiedad evita que la solución que se obtenga sea un promedio ponderado de otro par de soluciones factibles, lo que violaría una de las propiedades generales de las soluciones BF.) *Cualquier* conjunto de $n - 1$ arcos que no contenga ciclos no dirigidos formará un *árbol de expansión*. En consecuencia, cualquier conjunto completo de $n - 1$ arcos básicos es un árbol de expansión.

En consecuencia, las soluciones BF se pueden obtener "al resolver" árboles de expansión, como se resume a continuación.

Una **solución de un árbol de expansión** se obtiene de la siguiente manera:

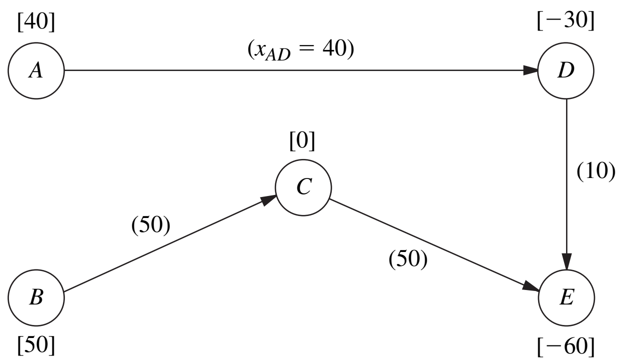
1. En el caso de los arcos que *no* están en el árbol de expansión (los arcos no básicos), se igualan a cero las variables correspondientes (x_{ij} o y_{ij}).
2. En el caso de los arcos que *sí* están en el árbol de expansión (los arcos básicos), se obtienen los valores de las variables correspondientes (x_{ij} o y_{ij}) resolviendo el sistema de ecuaciones lineales conformado por las restricciones de los nodos.

(El método símplex de redes obtiene los valores de la nueva solución BF a partir de la actual, de una manera mucho más eficiente, sin resolver este sistema de ecuaciones desde el principio.) Observe que este proceso de solución no considera ni las restricciones de no negatividad ni las restricciones de capacidad de los arcos para las variables básicas, por lo que la solución del árbol de expansión que se obtiene puede o no ser factible respecto de estas restricciones, lo cual conduce a la siguiente definición:

> Un **árbol de expansión factible** es un árbol de expansión cuya solución, a partir de las restricciones de nodos, también satisface todas las demás restricciones ($0 \leq x_{ij} \leq u_{ij}$ o $0 \leq y_{ij} \leq u_{ij}$).
>
> Con base en estas definiciones ahora es posible resumir la conclusión clave de la siguiente manera:
>
> El **teorema fundamental del método símplex de redes** sostiene que las soluciones básicas son *soluciones de árbol de expansión* (y a la inversa) y que las soluciones BF son soluciones de *árboles de expansión factibles* (y a la inversa).

Para ilustrar la aplicación de este teorema fundamental, considere la red que se muestra en la figura 9.16 que resulta al sustituir $x_{AB} = 10$ por $x_{AB} = 10 - y_{AB}$, en el ejemplo de la figura 9.12. En la figura 9.3*e* se muestra un árbol de expansión en esta red, en el que los arcos son $A \rightarrow D$, $D \rightarrow E$, $C \rightarrow E$ y $B \rightarrow C$. Con éstos como arcos básicos, a continuación se muestra el proceso de encontrar un árbol de expansión. En el lado izquierdo se encuentran las restricciones de los nodos dadas en la sección 9.6 después de sustituir x_{AB} por $10 - y_{AB}$, en donde las variables *básicas* aparecen en **negritas**. En el lado derecho, de arriba hacia abajo, se encuentran los pasos para establecer o calcular los valores de las variables.

$$
\begin{array}{lcrll}
 & & & \multicolumn{2}{l}{y_{AB} = 0,\ x_{AC} = 0,\ x_{ED} = 0} \\
\hline
-y_{AB} + x_{AC} + \boldsymbol{x_{AD}} & = & 40 & & x_{AD} = 40. \\
y_{AB} + \boldsymbol{x_{BC}} & = & 50 & & x_{BC} = 50. \\
- x_{AC} - \boldsymbol{x_{BC}} + \boldsymbol{x_{CE}} & = & 0 & \text{así} & x_{CE} = 50. \\
- \boldsymbol{x_{AD}} + \boldsymbol{x_{DE}} - x_{ED} & = & -30 & \text{así} & x_{DE} = 10. \\
- \boldsymbol{x_{CE}} - \boldsymbol{x_{DE}} + x_{ED} & = & -60 & \text{Redundante.} &
\end{array}
$$



■ **FIGURA 9.17**
Árbol de expansión factible inicial y su solución para el ejemplo.

Ya que los valores de todas estas variables satisfacen las restricciones de no negatividad y la única restricción relevante de capacidad de arco ($x_{CE} \leq 80$), el árbol de expansión es un *árbol de expansión factible*, por lo que se tiene una *solución BF*.

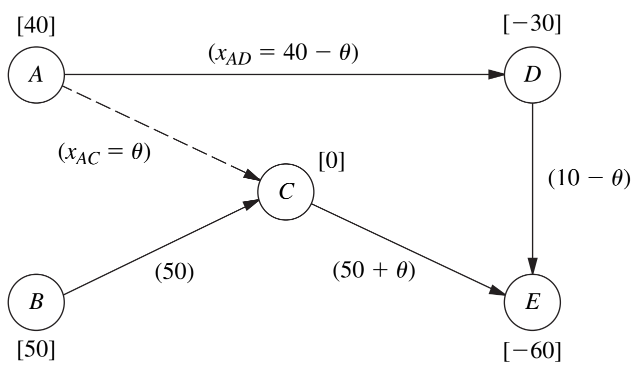
Se usará esta solución como la solución inicial BF para presentar el método símplex de redes. En la figura 9.17 se muestra su representación como una red, es decir, el árbol de expansión factible y su solución. En este caso, los números que se presentan junto a los arcos ahora representan *flujos* (valores de las x_{ij}) en lugar de los costos unitarios c_{ij} dados antes. (Como ayuda para distinguirlos, los flujos se pondrán entre paréntesis, no así los costos.)

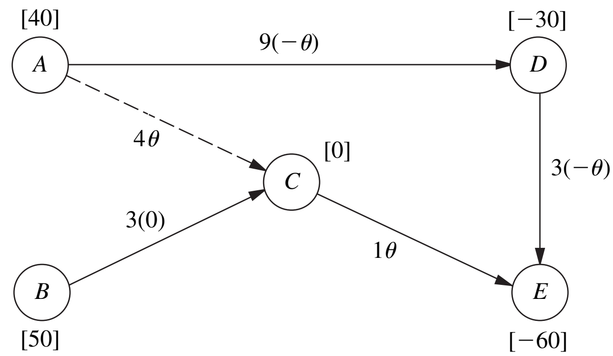
Selección de la variable básica entrante

Para comenzar una iteración del método símplex de redes, recuerde que el criterio del método símplex estándar para elegir la variable básica entrante es escoger la variable no básica que, al aumentar su valor, *mejore el valor de Z a la tasa más rápida*. Ahora se verá cómo se lleva a cabo este proceso sin tener una tabla símplex.

A manera de ilustración, considere la variable no básica x_{AC} de la solución BF inicial; es decir, el arco no básico $A \rightarrow C$. Aumentar x_{AC} desde cero hasta algún valor θ significa que debe agregarse a la red el arco $A \rightarrow C$ con flujo θ que se muestra en la figura 9.17. Si se agrega un arco no básico a un árbol de expansión, *siempre* se crea un *ciclo* no dirigido único que en este caso se observa en la figura 9.18 como *AC–CE–DE–AD*. En ella se muestra también el efecto de agregar el flujo θ al arco $A \rightarrow C$ sobre los otros flujos de la red. En particular, el flujo aumenta en θ en los arcos que tienen la misma dirección que $A \rightarrow C$ en el ciclo (el arco $C \rightarrow E$), mientras que el flujo neto disminuye en θ en otros arcos cuya dirección es opuesta a $A \rightarrow C$ en el ciclo (los arcos $D \rightarrow E$ y $A \rightarrow D$). En este último caso, el nuevo flujo, en efecto, cancela un flujo de θ en la dirección opuesta. El nuevo flujo θ no afecta los arcos que no están en el ciclo (el arco $B \rightarrow C$). (Verifique estas conclusiones al observar el efecto del cambio en x_{AC} sobre los valores de las otras variables en la solución que acaba de obtenerse para el árbol de expansión factible inicial.)

■ **FIGURA 9.18**
Efecto sobre los flujos al agregar el arco $A \rightarrow C$ con flujo θ al árbol de expansión factible inicial.





■ **FIGURA 9.19**
Efecto incremental sobre los costos al agregar el arco $A \rightarrow C$ con flujo θ al árbol de expansión factible inicial.

Ahora, ¿cuál es el efecto incremental sobre Z (costo total del flujo) si se agrega el flujo θ al arco $A \rightarrow C$? En la figura 9.19 se muestra la mayor parte de la respuesta, pues allí se proporciona el costo unitario multiplicado por el cambio en el flujo de cada arco de la figura 9.18. Por tanto, el incremento global en Z es

$$\begin{aligned}\Delta Z &= c_{AC}\theta + c_{CE}\theta + c_{DE}(-\theta) + c_{AD}(-\theta)\\ &= 4\theta + \theta - 3\theta - 9\theta\\ &= -7\theta.\end{aligned}$$

Si se establece $\theta = 1$ se obtiene la tasa de cambio de Z a medida que x_{AC} se incrementa; es decir,

$$\Delta Z = -7, \qquad \text{cuando } \theta = 1.$$

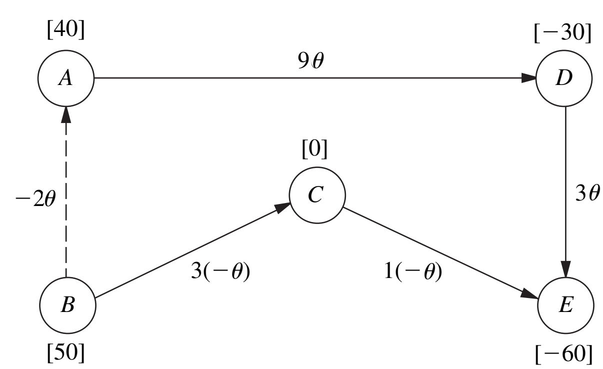
Como el objetivo es *minimizar* Z, esta tasa grande de disminución de Z al aumentar x_{AC} es muy deseable; por tanto, x_{AC} se convierte en un candidato de primer orden para ser la variable básica entrante.

Ahora es necesario realizar el mismo análisis sobre las otras variables no básicas antes de hacer la última selección de la variable básica entrante. Las únicas otras variables no básicas son y_{AB} y x_{ED}, correspondientes a los otros dos arcos no básicos, $B \rightarrow A$ y $E \rightarrow D$, en la figura 9.16.

En la figura 9.20 se muestra el efecto incremental sobre los costos al agregar el arco $B \rightarrow A$ con flujo θ al árbol de expansión factible inicial dado en la figura 9.17. Si se agrega este arco se crea un ciclo no dirigido BA–AD–DE–CE–BC, con lo que el flujo aumenta en θ para los arcos $A \rightarrow D$ y $D \rightarrow E$, pero disminuye en θ para los dos arcos en la dirección opuesta sobre este ciclo, $C \rightarrow E$ y $B \rightarrow C$. Estos incrementos de flujo θ y $-\theta$ son multiplicandos de los valores de los costos c_{ij} de la figura. Por tanto,

$$\begin{aligned}\Delta Z &= -2\theta + 9\theta + 3\theta + 1(-\theta) + 3(-\theta) = 6\theta\\ &= 6, \qquad \text{cuando } \theta = 1.\end{aligned}$$

■ **FIGURA 9.20**
Efecto incremental sobre los costos al agregar el arco $B \rightarrow A$ con flujo θ al árbol de expansión factible inicial.



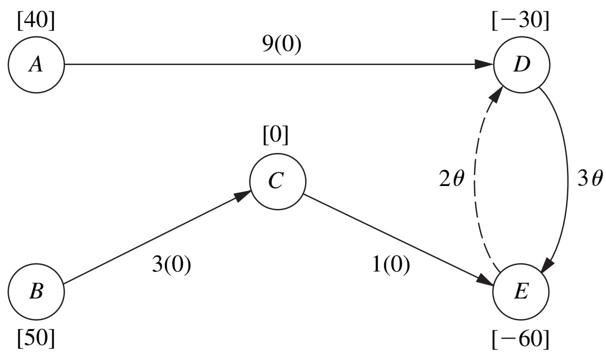


FIGURA 9.21
Efecto incremental sobre los costos al agregar el arco $E \to D$ con flujo θ al árbol de expansión factible inicial.

El hecho de que Z *aumente* en lugar de disminuir cuando y_{AB} (flujo a través del arco inverso $B \to A$) se incrementa, elimina esta variable como candidata a ser la variable básica entrante. (Recuerde que aumentar el valor de y_{AB} desde cero en realidad significa disminuir x_{AB}, es decir, el flujo a través del arco real $A \to B$ desde su cota superior de 10.)

En el caso del último arco no básico $E \to D$ se obtiene un resultado similar. Al agregar este arco con flujo de θ al árbol de expansión factible inicial, se crea un ciclo no dirigido *ED–DE* que se muestra en la figura 9.21, por lo que el flujo también se incrementa en θ en el arco $D \to E$, pero no se afecta ningún otro arco. Por tanto,

$$\begin{aligned}\Delta Z &= 2\theta + 3\theta = 5\theta\\ &= 5, \qquad \text{cuando } \theta = 1,\end{aligned}$$

por lo que x_{ED} se elimina como candidato a variable básica entrante.

Para resumir,

$$\Delta Z = \begin{cases} -7, & \text{si } \Delta x_{AC} = 1\\ \ \ 6, & \text{si } \Delta y_{AB} = 1\\ \ \ 5, & \text{si } \Delta x_{ED} = 1\end{cases}$$

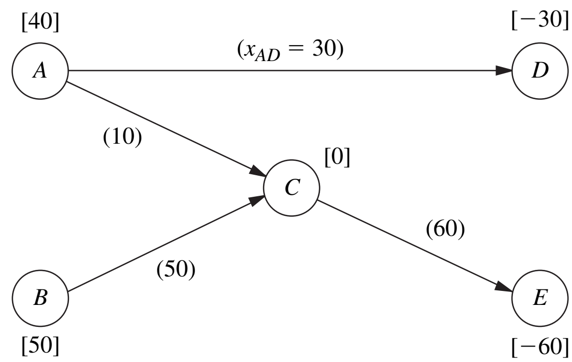
en consecuencia, el valor negativo de x_{AC} implica que x_{AC} se convierte en la variable básica entrante en la primera iteración. En caso de que haya más de una variable no básica con un valor *negativo* de ΔZ, se elige la que tiene el *mayor* valor absoluto. (Si no hay variables no básicas con un valor negativo de ΔZ, la solución BF actual es óptima.)

En lugar de identificar ciclos no dirigidos, etc., el método símplex de redes obtiene estos valores de ΔZ mediante un procedimiento algebraico que es mucho más eficiente, en especial en el caso de redes grandes. El procedimiento es análogo al que utiliza el método símplex de transporte (vea la sección 8.2) para encontrar los valores de las u_i y v_j con el fin de obtener el valor de $c_{ij} - u_i - v_j$ de cada variable no básica x_{ij}. No se describirá este procedimiento con más detalle, por lo que el lector deberá usar el método de los ciclos no dirigidos cuando deba resolver los problemas del final del capítulo.

Selección de la variable básica saliente y la próxima solución BF

Después de elegir la variable básica entrante, sólo se necesita un paso más para determinar al mismo tiempo la variable básica que sale y obtener la siguiente solución BF. En la primera iteración del ejemplo, la clave es la figura 9.18. Como x_{AC} es la variable básica entrante, se aumenta el flujo θ a través del arco $A \to C$ lo más posible, hasta que *una* de las variables básicas llegue a su cota inferior (0) o a su cota superior (u_{ij}). Para aquellos arcos cuyo flujo *aumenta* junto con θ en la figura 9.18 (los arcos $A \to C$ y $C \to E$), sólo es necesario considerar las cotas *superiores* ($u_{AC} = \infty$ y $u_{CE} = 80$):

$$\begin{aligned} x_{AC} &= \theta \leq \infty.\\ x_{CE} &= 50 + \theta \leq 80, \qquad \text{entonces} \qquad \theta \leq 30.\end{aligned}$$



■ **FIGURA 9.22**
Segundo árbol de expansión factible y su solución para el ejemplo.

En el caso de los arcos cuyo flujo *disminuye* con θ (los arcos $D \to E$ y $A \to D$), sólo tiene que considerarse la cota inferior de 0:

$$x_{DE} = 10 - \theta \geq 0, \qquad \text{entonces} \qquad \theta \leq 10.$$
$$x_{AD} = 40 - \theta \geq 0, \qquad \text{entonces} \qquad \theta \leq 40.$$

Los arcos cuyo flujo no cambia con θ —por ejemplo, aquellos que no forman parte del ciclo no dirigido—, que en este caso es sólo el arco B $\to$ C de la figura 9.18, se pueden ignorar puesto que no llegarán a ninguna de sus cotas cuando θ aumenta.

En el caso de los cinco arcos de la figura 9.18, la conclusión es que x_{DE} debe ser la variable básica que sale, puesto que alcanza una cota para el valor más pequeño de θ (10). Con $\theta = 10$ en esta figura, se obtienen los flujos a través de los arcos básicos en la siguiente solución BF:

$$x_{AC} = \theta = 10,$$
$$x_{CE} = 50 + \theta = 60,$$
$$x_{AD} = 40 - \theta = 30,$$
$$x_{BC} = 50.$$

El árbol de expansión factible correspondiente se muestra en la figura 9.22.

Si la variable básica que sale hubiera alcanzado su cota superior, entonces en este punto se habrían necesitado los ajustes que se presentaron para la técnica de la cota superior, como se verá durante las siguientes dos iteraciones. Como se alcanzó la cota inferior de 0, no es necesario hacer más.

Terminación del ejemplo. En las dos iteraciones que faltan para llegar a la solución óptima, la atención se centrará en algunas características de la técnica de la cota superior que éstas ilustran. El patrón para encontrar la variable básica entrante, la que sale y la siguiente solución BF será muy parecido al descrito en la primera iteración, así que sólo se presentará un resumen de estos pasos.

Iteración 2: A partir del árbol de expansión factible que se muestra en la figura 9.22 y con referencia a la figura 9.16 para los costos unitarios c_{ij}, se llega a los cálculos para seleccionar la variable básica entrante que se presentó en la tabla 9.4. La segunda columna identifica el ciclo no dirigido único que se crea al agregar el arco no básico de la primera columna al árbol de expansión, y la tercera columna muestra el efecto incremental sobre los costos debido a los cambios en los flujos de este ciclo causados por agregar un flujo de $\theta = 1$ al arco no básico. El arco $E \to D$ tiene el mayor valor negativo (en términos absolutos) de ΔZ, de manera que x_{ED} es la variable básica entrante.

■ **TABLA 9.4** **Cálculos para elegir la variable básica entrante en la iteración 2**

Arco no básico	Ciclo creado	ΔZ cuando $\theta = 1$	
$B \to A$	*BA–AC–BC*	$-2 + 4 - 3 = -1$	
$D \to E$	*DE–CE–AC–AD*	$3 - 1 - 4 + 9 = 7$	
$E \to D$	*ED–AD–AC–CE*	$2 - 9 + 4 + 1 = -2$	← Mínimo

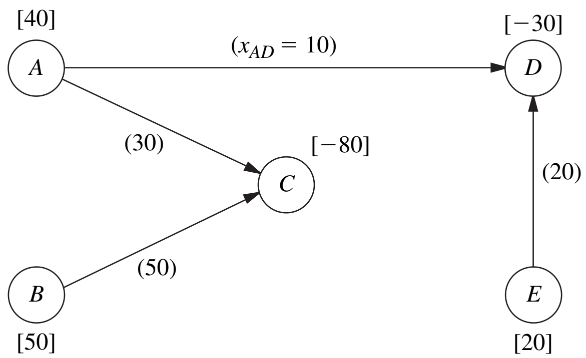


FIGURA 9.23
Tercer árbol de expansión factible y su solución para el ejemplo.

Ahora el flujo θ en el arco $E \rightarrow D$ se hace tan grande como sea posible, sin violar las siguientes cotas sobre los flujos:

$$
\begin{array}{lll}
x_{ED} = \theta \le u_{ED} = \infty, & \text{así} & \theta \le \infty. \\
x_{AD} = 30 - \theta \ge 0, & \text{así} & \theta \le 30. \\
x_{AC} = 10 + \theta \le u_{AC} = \infty, & \text{así} & \theta \le \infty. \\
x_{CE} = 60 + \theta \le u_{CE} = 80, & \text{así} & \theta \le 20. \quad \leftarrow \text{Mínimo}
\end{array}
$$

Como x_{CE} impone la menor cota superior (20) sobre θ, x_{CE} se convierte en la variable básica que sale. Con $\theta = 20$ en las expresiones anteriores, para x_{ED}, x_{AD} y x_{AC}, se obtiene el flujo a través de los arcos básicos de la siguiente solución BF (con $x_{BC} = 50$ sin ser afectado por θ), como se muestra en la figura 9.23.

De interés especial en este punto es que la variable básica que sale (x_{CE}) se obtuvo a partir de la variable que alcanzó su cota superior (80). Por tanto, mediante la técnica de la cota superior, x_{CE} se sustituye por $80 - y_{CE}$, donde $y_{CE} = 0$ es la nueva variable no básica. Al mismo tiempo, el arco original $C \rightarrow E$ con $c_{CE} = 1$ y $u_{CE} = 80$ se sustituye por el arco invertido $E \rightarrow C$ con $c_{EC} = -1$ y $u_{EC} = 80$. Los valores de b_E y b_C también se ajustan al agregar 80 a b_E y restar 80 de b_C. La red ajustada que resulta se muestra en la figura 9.24, donde los arcos no básicos se presentan mediante líneas punteadas, mientras que los números junto a todos los arcos son los costos unitarios.

Iteración 3: Para iniciar la siguiente iteración se usan las figuras 9.23 y 9.24. En la tabla 9.5 se muestran los cálculos que conducen a la elección de y_{AB} (inverso del arco $B \rightarrow A$) como la variable básica entrante. Después se agrega tanto flujo θ como sea posible al arco $B \rightarrow A$, sin que violen las siguientes cotas de los flujos:

$$
\begin{array}{lll}
y_{AB} = \theta \le u_{BA} = 10, & \text{así} & \theta \le 10. \quad \leftarrow \text{Mínimo} \\
x_{AC} = 30 + \theta \le u_{AC} = \infty, & \text{así} & \theta \le \infty. \\
x_{BC} = 50 - \theta \ge 0, & \text{así} & \theta \le 50.
\end{array}
$$

La variable y_{AB} impone la cota superior más pequeña (10) sobre θ, por lo que se convierte en la variable básica que sale. Si se hace $\theta = 10$ en estas expresiones para x_{AC} y x_{BC} —junto con los valores

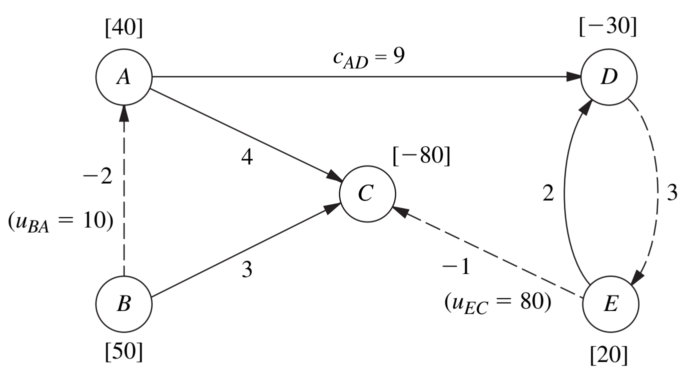


FIGURA 9.24
Red ajustada con costos unitarios al terminar la iteración 2.

TABLA 9.5 Cálculos para elegir la variable básica entrante en la iteración 3

Arco no básico	Ciclo creado	ΔZ cuando $\theta = 1$	
$B \to A$	BA–AC–BC	$-2 + 4 - 3 = -1$	← Mínimo
$D \to E$	DE–ED	$3 + 2 = 5$	
$E \to C$	EC–AC–AD–ED	$-1 - 4 + 9 - 2 = 2$	

que no cambian de $x_{AC} = 10$ y $x_{ED} = 20$— se obtiene la siguiente solución básica factible que se muestra en la figura 9.25.

Igual que en la iteración 2, la variable básica que sale (y_{AB}) se obtuvo con la variable que alcanzó su cota superior. Además, existen otros dos puntos de interés especial respecto de esta elección particular. Uno es que la variable básica *entrante* y_{AB} también se convierte en la variable básica que *sale* en la misma iteración. En ocasiones, este fenómeno ocurre con la técnica de la cota superior cuando al aumentar el valor de la variable básica entrante, ésta alcanza su cota superior antes de que lo haga cualquier otra de las variables básicas con su cota.

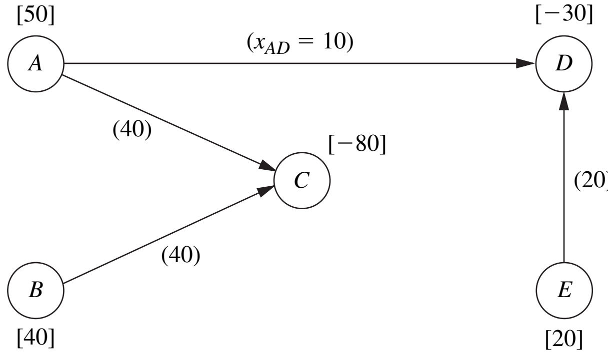
El otro punto de interés es que el arco $B \to A$ que ahora es necesario cambiar por un arco *inverso* $A \to B$ —porque la variable básica que sale alcanza una cota superior— ya es un arco inverso. Esto no ocasiona ningún problema puesto que el arco inverso de un arco inverso es simplemente el arco real *original*. Por tanto, el arco $B \to A$ (con $c_{BA} = -2$ y $u_{BA} = 10$) de la figura 9.24 ahora se sustituye por el arco $A \to B$ (con $c_{AB} = 2$ y $u_{AB} = 10$), que es el arco entre los nodos A y B de la red original de la figura 9.12 y se cambia un flujo generado de 10 del nodo B ($b_B = 50 \to 40$) al nodo A ($b_A = 40 \to 50$). Al mismo tiempo, la variable $y_{AB} = 10$ se sustituye por $10 - x_{AB}$, con $x_{AB} = 0$ como la nueva variable no básica. En la figura 9.26 se muestra la red ajustada que resulta.

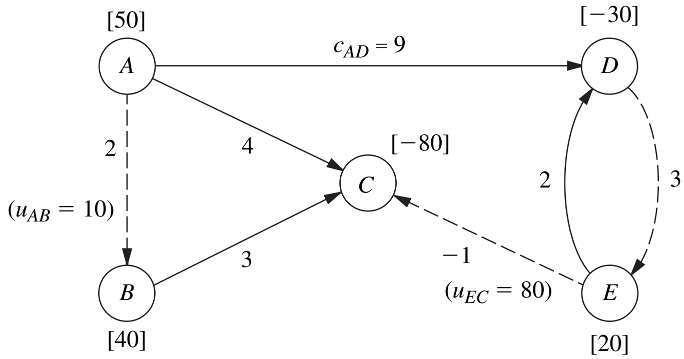
Prueba de optimalidad: En este punto, el algoritmo intentará usar las figuras 9.25 y 9.26 para encontrar la siguiente variable básica entrante con los cálculos normales que se muestran en la tabla 9.6. Sin embargo, *ninguno* de los arcos no básicos da un valor *negativo* de ΔZ, de manera que *no puede* lograrse una mejora del valor de Z al introducir un flujo a través de alguno de ellos. Esto significa que la solución básica factible actual que se muestra en la figura 9.25 *pasa* la prueba de optimalidad y el algoritmo se detiene.

Para identificar los flujos a través de los arcos reales, y no de los inversos en el caso de esta solución óptima, la red ajustada actual (figura 9.26) se debe comparar con la red original (figura 9.12). Observe que cada uno de los arcos tiene la misma dirección en las dos redes, con la única excepción del arco entre los nodos C y E. Esto significa que el único arco inverso en la figura 9.26 es el arco $E \to C$, donde el flujo está dado por la variable y_{CE}. En consecuencia, se calcula $x_{CE} = u_{CE} - y_{CE} = 80 - y_{CE}$. El arco $E \to C$ resulta ser un arco no básico, entonces, $y_{CE} = 0$ y $x_{CE} = 80$ es el flujo a través del arco real $C \to E$. Todos los demás flujos a través de los arcos reales son los flujos dados en la figura 9.25. Por tanto, la solución óptima es la que se muestra en la figura 9.27.

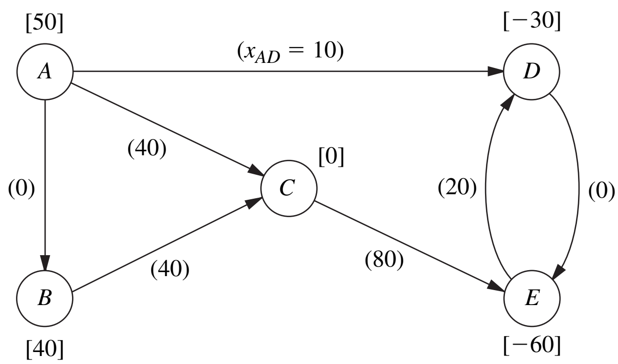
Se proporciona **otro ejemplo completo** de la aplicación del método símplex de redes en la demostración que se presenta en el área de análisis de redes (*Network Analysis Area*) del OR Tutor. En la sección Worked Examples del sitio en internet de este libro se incluye un **ejemplo adicional**. También el IOR Tutorial incluye una rutina interactiva del método símplex de redes.

FIGURA 9.25
Cuarto árbol de expansión factible (y final) y su solución para el ejemplo.





■ **FIGURA 9.26**
Red ajustada con costos unitarios al terminar la iteración 3.



■ **FIGURA 9.27**
Patrón de flujo óptimo en la red original del ejemplo de Distribution Unlimited Co.

■ **TABLA 9.6 Cálculos para elegir la variable básica entrante en la iteración 3**

Arco no básico	Ciclo creado	ΔZ cuando $\theta = 1$
$A \to B$	AB–BC–AC	$2 + 3 - 4 = 1$
$D \to E$	DE–ED	$3 + 2 = 5$
$E \to C$	EC–AC–AD–ED	$-1 - 4 + 9 - 2 = 2$

9.8 MODELO DE REDES PARA OPTIMIZAR LOS TRUEQUES ENTRE TIEMPO Y COSTO DE UN PROYECTO

Las redes proporcionan una forma natural de desplegar de manera gráfica el flujo de actividades de un proyecto a gran escala o de un proyecto de investigación y desarrollo. Por tanto, una de las aplicaciones más importantes de la teoría de redes es ayudar en la gestión de tales proyectos.

Al final de la década de 1950, se desarrollaron dos técnicas de IO basadas en redes —**PERT** (técnica de evaluación y revisión de programas o *program evaluation and review technique*) y **CPM** (método de la ruta crítica o *critical path method*)— para ayudar a los directores de proyectos a cumplir con sus responsabilidades. Estas técnicas fueron diseñadas para auxiliar en la planeación y la coordinación de las distintas actividades de un proyecto, desarrollar un calendario realista para éste y luego supervisar el avance del proyecto después de que éste se puso en marcha. Con el paso de los años, las mejores características de estas dos técnicas han tendido a fusionarse en algo que ahora se conoce de manera común como técnica PERT/CPM. Este enfoque de redes en la administración de proyectos se usa ampliamente en la actualidad.

Uno de los capítulos complementarios del sitio en internet que acompaña a este libro, el capítulo 22 (Administración de proyectos con PERT/CPM o *Project Management with PERT/CPM*), proporciona una descripción completa de las diferentes características de PERT/CPM. Ahora se

concentrará la atención en una de estas características por dos razones. Primera, que es un modelo de optimización de redes, por lo cual se ajusta al tema del presente capítulo. Segunda, ilustra el tipo de aplicaciones importantes que tal modelo puede tener.

La característica que se destacará se conoce como *método CPM de trueques entre tiempo y costo* porque fue una parte importante de la técnica original del CPM. Aborda el siguiente problema de un proyecto que debe completarse para una fecha límite específica. Suponga que esta fecha límite no se cumpliría si todas las actividades se realizan de la manera normal, pero que existen varias formas de cumplir con ella mediante la inversión de más dinero para acelerar algunas de las actividades. ¿Cuál es el plan óptimo para acelerar algunas actividades de forma que se minimice el costo total de completar el proyecto dentro del plazo establecido?

El enfoque general comienza con el uso de una red para desplegar las diferentes actividades y el orden en que éstas deben realizarse. Después se formula un modelo de optimización que puede resolverse por medio del análisis de costo marginal o de programación lineal. Igual que con los otros modelos de optimización de redes que se consideraron antes en este capítulo, la estructura especial del problema lo hace relativamente fácil de resolver con eficiencia.

Este enfoque se ejemplifica a continuación mediante el uso del mismo ejemplo prototipo que se estudia a lo largo del capítulo 22.

Ejemplo prototípico: Proyecto de Reliable Construction Co.

La compañía RELIABLE CONSTRUCTION acaba de ganar una licitación de 5.4 millones de dólares para construir una nueva planta para un fabricante importante. El dueño de la empresa necesita que la planta esté en operación en 40 semanas.

Reliable ha asignado este proyecto a su mejor director de construcción, David Perty, para asegurar que se complete a tiempo. El señor Perty debe hacer los arreglos para que cierto número de brigadas realicen las distintas actividades de construcción en diferentes tiempos. En la tabla 9.7 se muestra la lista de actividades. La tercera columna proporciona información importante para la programación de las brigadas.

> Para una actividad dada, sus **predecesores inmediatos** —que se muestra en la tercera columna de la tabla 9.7— son aquellas actividades que no deben terminarse después del tiempo de inicio de la actividad en cuestión. (De manera similar, la actividad dada se llama **sucesor inmediato** de cada uno de sus predecesores inmediatos.)

Por ejemplo, los primeros elementos en esta columna indican que

1. La excavación no requiere que se cumpla ninguna otra actividad.
2. La excavación debe terminar antes de comenzar a colocar los cimientos.
3. Los cimientos deben quedar colocados por completo antes de iniciar la obra negra, etcétera.

TABLA 9.7 Lista de actividades para el proyecto de la Reliable Construction Co.

Actividad	Descripción de la actividad	Predecesores inmediatos	Duración estimada
A	Excavación	—	2 semanas
B	Colocar los cimientos	*A*	4 semanas
C	Levantar paredes	*B*	10 semanas
D	Colocar el techo	*C*	6 semanas
E	Instalar la plomería exterior	*C*	4 semanas
F	Instalar la plomería interior	*E*	5 semanas
G	Aplanados exteriores	*D*	7 semanas
H	Pintura exterior	*E, G*	9 semanas
I	Instalar el cableado eléctrico	*C*	7 semanas
J	Aplanados interiores	*F, I*	8 semanas
K	Colocar pisos	*J*	4 semanas
L	Pintura interior	*J*	5 semanas
M	Colocar accesorios exteriores	*H*	2 semanas
N	Colocar accesorios interiores	*K, L*	6 semanas

Cuando una actividad dada tiene *más de un predecesor inmediato*, todos deben estar terminados antes de iniciarla.

Para programar las actividades, el señor Perty consulta a cada uno de los supervisores de brigada para desarrollar una estimación del tiempo que debe tomar cada actividad si se realiza de manera normal. Estas estimaciones se presentan en la columna de la derecha de la tabla 9.7.

La suma de estos tiempos da un gran total de 79 semanas, que es mucho más que la fecha de entrega del proyecto. Por fortuna, algunas actividades se pueden realizar en paralelo, lo que reducirá de modo sustancial el tiempo de terminación. En seguida se verá cómo se puede desplegar el proyecto en forma gráfica para visualizar de mejor manera el flujo de las actividades y para determinar el tiempo total que se requiere para completar el proyecto si no ocurren retrasos.

Ya se ha visto en este capítulo cómo se pueden representar *redes* valiosas para ayudar a analizar muchos tipos de problemas. De manera muy parecida, las redes tienen un papel clave en el manejo de proyectos, pues permiten mostrar las relaciones entre las actividades y colocar el plan en una buena perspectiva global. También son útiles para analizar el proyecto.

Redes de proyectos

Una red que se utiliza para representar un proyecto se llama **red de proyecto**, que consiste en cierto número de *nodos* —que se muestran como pequeños círculos o rectángulos— y *arcos* —como flechas— que van de algún nodo a otro.

Como indica la tabla 9.7, se requiere de tres tipos de datos para describir un proyecto.

1. Información de la actividad: se desglosa el proyecto en sus *actividades* individuales (al nivel de detalle deseado).
2. Relaciones de precedencia: se identifican *los predecesores inmediatos* de cada actividad.
3. Datos de tiempo: se estima la *duración* de cada actividad.

La red de proyecto debe contener toda esta información. Se dispone de dos tipos alternativos de redes de proyecto para llevar a cabo esta tarea.

Un tipo es la red de proyecto de **actividades en los arcos** (**AOA**, *activity on arc*), donde cada actividad está representada por un *arco*. Un nodo se usa para separar una actividad (arco que sale) de cada uno de sus predecesores inmediatos (arcos que llegan). Entonces, la secuencia de arcos muestra las relaciones de precedencia entre las actividades.

El segundo tipo es la red de proyecto de **actividades en los nodos** (**AON**, *activity on node*), donde cada actividad se representa por un *nodo*. Los arcos se usan sólo para indicar las relaciones de precedencia entre las actividades. En particular, el nodo de cada actividad con los predecesores inmediatos tiene un arco que llega desde cada uno de estos predecesores.

Las versiones originales de PERT y CPM usaban las redes de proyecto AOA, por lo que se convirtieron en el tipo convencional durante algunos años. Sin embargo, las redes de proyecto AON tienen algunas ventajas importantes sobre las redes AOA para comunicar exactamente la misma información.

1. Es mucho más sencillo construir las redes de proyecto AON que las redes AOA.
2. Es más fácil que los usuarios no experimentados, categoría que incluye a muchos administradores, entiendan más las redes de proyecto AON que las AOA.
3. Es más sencillo revisar las redes de proyecto AON que las AOA cuando hay cambios en el proyecto.

Por estas razones, las redes AON son cada vez más populares entre quienes las aplican. En apariencia, es posible que se conviertan en el tipo de red de uso convencional. Por tanto, el estudio se centrará sólo en las redes de proyecto AON, y se eliminará el adjetivo que las etiqueta.

En la figura 9.28 se muestra la red de proyecto para el contrato de Reliable.[2] Con referencia a la tercera columna de la tabla 9.7, se observa que existe un arco que llega a cada actividad desde cada uno de sus predecesores inmediatos. Como la actividad *A* no tiene predecesores, se tiene un nodo que va del nodo de inicio a esta actividad. De manera similar, como las actividades *M* y *N* no tienen sucesores inmediatos, los arcos que salen de estas actividades llegan al nodo de termina-

[2] Aunque con frecuencia las redes de proyecto se dibujan de izquierda a derecha, ésta se trazará de arriba hacia abajo para que se ajuste de mejor forma a la página impresa.

ción. De esta forma, la red de proyecto muestra, en un vistazo, todas las relaciones de precedencia entre todas las actividades (más el inicio y la terminación del proyecto). Según la columna de la derecha de la tabla 9.7, el número al lado del nodo de cada actividad registra la duración estimada (en semanas) de esa actividad.

La ruta crítica

¿Cuánto tiempo se debe emplear en el proyecto? Se observó antes que al sumar las duraciones de todas las actividades se obtiene un gran total de 79 semanas. Sin embargo, ésta no es la respuesta a la pregunta porque algunas actividades se pueden realizar (más o menos) al mismo tiempo.

En consecuencia, lo relevante es la *longitud* de cada *trayectoria* a través de la red.

> Una **trayectoria** a través de una red de proyecto es una de las rutas que siguen los arcos desde el nodo INICIO hasta el nodo TERMINACIÓN. La **longitud** de una trayectoria es la *suma* de las *duraciones* (estimadas) de las actividades en la trayectoria.

Las seis trayectorias que cruzan la red de la figura 9.28 se presentaron en la tabla 9.8, junto con los cálculos de sus longitudes. Las longitudes de las trayectorias van de 31 a 44 semanas en el caso de la trayectoria más larga (la cuarta de la tabla).

Así, dadas estas longitudes, ¿cuál debe ser la **duración del proyecto** (estimada), es decir, el tiempo total que se requiere para terminarlo? A continuación se analizará este tema.

Como las actividades en cualquier trayectoria específica deben realizarse una después de otra sin superposiciones, la duración del proyecto no puede ser *menor* que la longitud de la trayectoria. Sin embargo, la duración del proyecto podría ser *mayor* si alguna actividad en la trayectoria con predecesores inmediatos múltiples tuviera que esperar a que terminara un predecesor inmediato que *no* estuviera en la trayectoria más tiempo que los que se encontraran en ella. Por ejemplo, consi-

■ **FIGURA 9.28**
Red de proyecto del contrato de Reliable Construction Co.

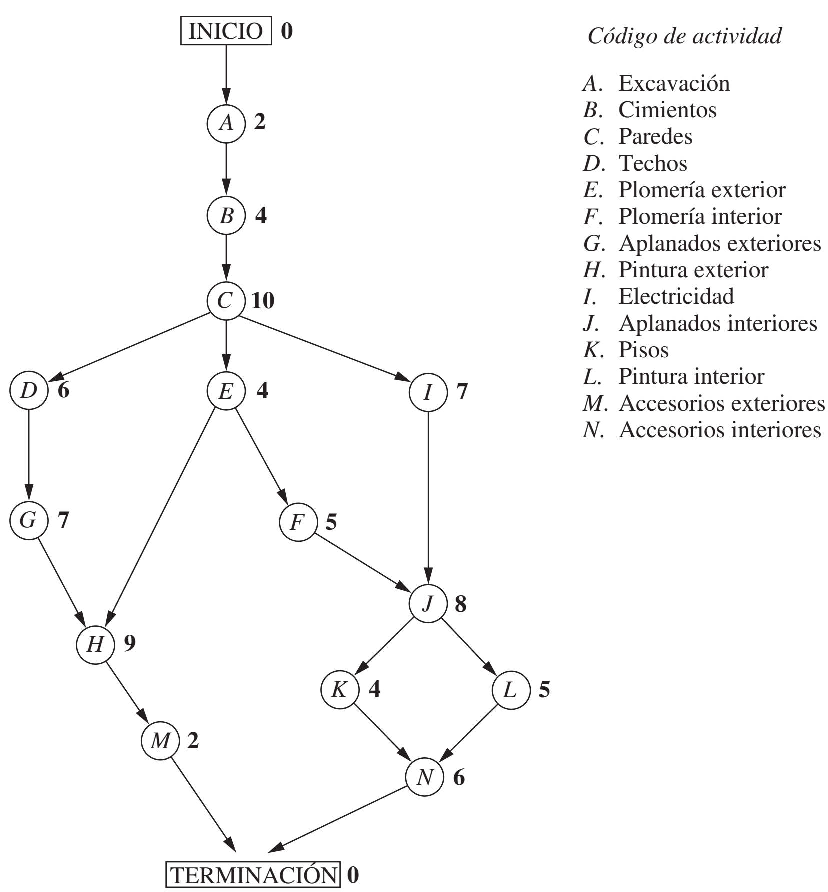


TABLA 9.8 Trayectorias y sus longitudes a través de la red de proyecto de Reliable

Trayectoria	Longitud
INICIO →A→B→C→D→G→H→M→ TERMINACIÓN	2 + 4 + 10 + 6 + 7 + 9 + 2 = 40 semanas
INICIO →A→B→C→E→H→M→ TERMINACIÓN	2 + 4 + 10 + 4 + 9 + 2 = 31 semanas
INICIO →A→B→C→E→F→J→K→N→ TERMINACIÓN	2 + 4 + 10 + 4 + 5 + 8 + 4 + 6 = 43 semanas
INICIO →A→B→C→E→F→J→L→N→ TERMINACIÓN	2 + 4 + 10 + 4 + 5 + 8 + 5 + 6 = 44 semanas
INICIO →A→B→C→I→J→K→N→ TERMINACIÓN	2 + 4 + 10 + 7 + 8 + 4 + 6 = 41 semanas
INICIO →A→B→C→I→J→L→N→ TERMINACIÓN	2 + 4 + 10 + 7 + 8 + 5 + 6 = 42 semanas

dere la segunda trayectoria de la tabla 9.8 y observe la actividad H. Ésta tiene dos predecesores: uno (la actividad G) *no* está en la trayectoria, y el otro (la actividad E), sí lo está. Una vez terminada la actividad C, sólo se requieren cuatro semanas más para la actividad E, pero se necesitan 13 semanas para que termine la actividad D y después la actividad G. Por tanto, la duración del proyecto debe ser mucho mayor que la longitud de la segunda trayectoria de la tabla.

No obstante, la duración del proyecto no será mayor que una trayectoria en particular. Ésta es la *ruta más larga* a través de la red. Las actividades en esta ruta se pueden realizar en secuencia ininterrumpida. (De otra manera, no sería la ruta más larga.) Por tanto, el tiempo que se requiere para llegar al nodo de TERMINACIÓN es igual a la longitud de esta trayectoria. Más aún, las rutas más cortas no pueden llegar a la TERMINACIÓN después de esto.

La conclusión clave es la siguiente:

> La *duración del proyecto* (estimada) es igual a la *longitud de la ruta más larga* a través de la red de proyecto. Esta trayectoria más larga se llama **ruta crítica**.[3] (Si varias trayectorias son iguales, todas son rutas críticas.)

Así, en el caso del proyecto de Reliable Construction Co., se tiene

> Ruta crítica: INICIO →A→B→C→E→F→J→L→N→ TERMINACIÓN
> Duración (estimada) del proyecto = 44 semanas.

Por tanto, si no ocurren retrasos, el tiempo total que se requiere para terminar el proyecto debe ser de alrededor de 44 semanas. Aún más, las actividades en esta ruta crítica son las actividades cuello de botella críticas en las que debe evitarse cualquier demora en su terminación para prevenir que se retrase la conclusión del proyecto. Esta información es invaluable para el señor Perty, puesto que ahora sabe que debe dedicar la mayor parte de su atención a que estas actividades se realicen a tiempo para cumplir con el programa de todo el proyecto. Además, si decide reducir la duración del proyecto —recuerde que el límite para terminar es de 40 semanas—, ésas son las principales actividades en las cuales deben hacerse cambios para reducir sus duraciones.

Ahora, el señor Perty debe determinar de manera específica cuáles actividades deben reducir sus duraciones, y en qué medida, con el propósito de cumplir con el límite de 40 semanas en la forma menos cara. El señor Perty recuerda que el CPM proporciona un procedimiento excelente para investigar esos *trueques entre tiempo y costo*, de modo que nuevamente usará este enfoque para estudiar esta pregunta.

Primero se presentarán algunos antecedentes.

Trueques entre tiempo y costo de actividades individuales

El primer concepto clave de este enfoque es el de *quiebre*.

> El **quiebre de una actividad** se refiere a tomar medidas especiales y costosas para reducir la duración de una actividad a menos de su estimación normal. Estas medidas especiales

[3] Aunque la tabla 9.8 ilustra cómo se puede usar la enumeración de rutas y la de sus longitudes para encontrar la ruta crítica en proyectos pequeños, en el capítulo 22 se describe la forma en que el PERT/CPM normalmente usa un procedimiento mucho más eficiente para obtener una información variada y útil, la cual incluye la ruta crítica.

incluyen el uso de horas extra, contratación temporal de ayuda, uso de materiales especiales que ahorran tiempo, obtención de equipo especial, etc. El **quiebre de un proyecto** se refiere a acelerar cierto número de actividades para reducir la duración del proyecto a menos de su estimación normal.

El **método CPM de trueques entre tiempo y costo** se ocupa de determinar cuánto acelerar (si se hace) cada actividad para reducir la duración prevista del proyecto a una estimación deseada.

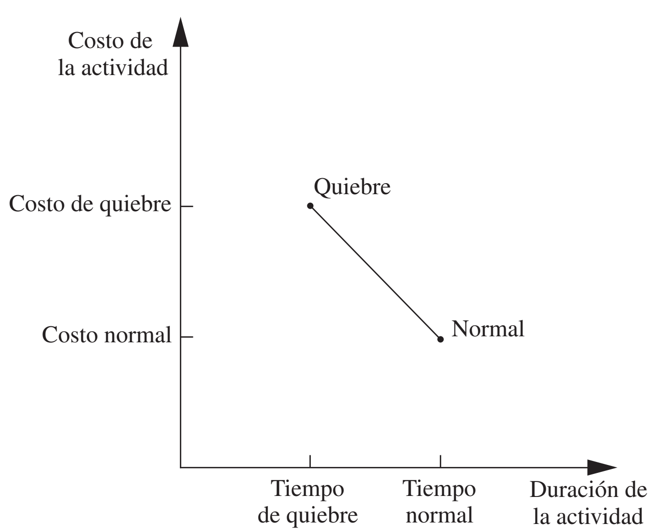
Los datos necesarios para determinar cuánto acelerar una actividad específica están dados por la *gráfica de tiempo-costo* de la actividad. En la figura 9.29 se muestra una gráfica típica. Observe los dos puntos importantes en esta gráfica llamados *normal* y de *quiebre*.

El **punto normal** de una gráfica de tiempo-costo de una actividad muestra el tiempo (duración) y el costo de la actividad cuando se realiza de manera normal. El **punto de quiebre** muestra el tiempo y el costo cuando la actividad se *acelera por completo*; es decir, se fuerza sin límite de costo para reducir su duración todo lo posible. Como una aproximación, el CPM supone que estos tiempos y costos se pueden predecir de manera confiable sin incertidumbre significativa.

En el caso de la mayoría de las aplicaciones, se supone que una actividad con *quiebre parcial* en cualquier nivel resultará en una combinación de tiempo y costo que está en algún punto del segmento de recta entre estos dos puntos.[4] (Por ejemplo, este supuesto sostiene que la *mitad* de un quiebre completo se encuentra en un punto de este segmento a la mitad entre los puntos normal y de quiebre.) Esta aproximación simplificada facilita la recolección de datos para estimar el tiempo y costo en sólo dos situaciones: *condiciones normales* (para obtener el punto normal) y *quiebre completo* (para obtener el punto de quiebre).

Con base en este enfoque, el señor Perty ha pedido a su personal y supervisores que desarrollen estos datos para cada una de las actividades del proyecto de Reliable. Por ejemplo, el supervisor de la brigada responsable de colocar los aplanados indica que si agregan dos empleados temporales y usa tiempo extra podrá reducir la duración de esta actividad de 8 a 6 semanas, que es el mínimo posible. Entonces, el personal del señor Perty estima el costo de un quiebre completo de esta forma para esta actividad, comparada con la siguiente programación normal de 8 semanas, como se muestra en seguida.

FIGURA 9.29
Gráfica tiempo-costo típica de una actividad.



[4] Éste es un supuesto conveniente, pero a menudo se trata sólo de una aproximación burda dado que los supuestos subyacentes de proporcionalidad y divisibilidad pueden no cumplirse por completo. Si la gráfica verdadera de tiempo-costo es convexa, aún se puede usar la programación lineal mediante el empleo de una aproximación lineal, para después aplicar la técnica de programación separable descrita en la sección 12.8.

Actividad J (colocar aplanados):

Punto normal: tiempo = 8 semanas, costo = \$430 000.
Tiempo de quiebre: tiempo = 6 semanas, costo = \$490 000.
Reducción máxima de tiempo = 8 − 6 = 2 semanas.

$$\text{Costo de quiebre por semana ahorrada} = \frac{\$490\,000 - \$430\,000}{2}$$
$$= \$30\,000.$$

Después de investigar de la misma manera el trueque entre tiempo y costo de cada una de las actividades restantes, en la tabla 9.9 se proporcionan los datos que se obtuvieron para todas las actividades.

¿Qué actividades deben acelerarse?

Cuando se suman las columnas de *costo normal* y *costo de quiebre* de la tabla 9.9 se obtiene:

Suma de costos normales = \$4.55 millones,
Suma de costos de quiebre = \$6.15 millones.

Recuerde que la compañía recibirá 5.4 millones de dólares por realizar este proyecto. Este pago debe cubrir algunos *costos generales* además de los costos de las actividades enumeradas en la tabla, lo mismo que proporcionar una ganancia razonable para la compañía. Al desarrollar el presupuesto (ganador) de 5.4 millones de dólares, la administración de Reliable pensó que esta cantidad proporcionaría una ganancia razonable siempre que el costo total de las actividades se pudiera mantener cercano al nivel de 4.55 millones de dólares. El señor Perty entiende muy bien que ahora es su responsabilidad mantener el proyecto lo más posible dentro del presupuesto y del tiempo estipulado.

Como se vio en la tabla 9.8, si todas las actividades se realizan de manera normal, la duración prevista del proyecto sería de 44 semanas (si pueden evitarse los retrasos). Si *todas* las actividades tuvieran un *quiebre completo*, un cálculo similar encontraría que esta duración se reduciría a sólo 28 semanas. Sin embargo, el costo de hacerlo sería prohibitivo (6.15 millones de dólares). Es claro que la aceleración completa de todas las actividades no es una opción que se deba considerar.

De cualquier modo, el señor Perty desea investigar la posibilidad de un quiebre parcial o completo de sólo algunas actividades para reducir la duración del proyecto a 40 semanas.

TABLA 9.9 Datos de trueques entre tiempo y costo de las actividades del proyecto de Reliable

Actividad	Tiempo		Costo		Máxima reducción en tiempo	Costo de quiebre por semana ahorrada
	Normal	Quiebre	Normal	Quiebre		
A	2 semanas	1 semana	\$180 000	\$ 280 000	1 semana	\$100 000
B	4 semanas	2 semanas	\$320 000	\$ 420 000	2 semanas	\$ 50 000
C	10 semanas	7 semanas	\$620 000	\$ 860 000	3 semanas	\$ 80 000
D	6 semanas	4 semanas	\$260 000	\$ 340 000	2 semanas	\$ 40 000
E	4 semanas	3 semanas	\$410 000	\$ 570 000	1 semana	\$160 000
F	5 semanas	3 semanas	\$180 000	\$ 260 000	2 semanas	\$ 40 000
G	7 semanas	4 semanas	\$900 000	\$1 020 000	3 semanas	\$ 40 000
H	9 semanas	6 semanas	\$200 000	\$ 380 000	3 semanas	\$ 60 000
I	7 semanas	5 semanas	\$210 000	\$ 270 000	2 semanas	\$ 30 000
J	8 semanas	6 semanas	\$430 000	\$ 490 000	2 semanas	\$ 30 000
K	4 semanas	3 semanas	\$160 000	\$ 200 000	1 semana	\$ 40 000
L	5 semanas	3 semanas	\$250 000	\$ 350 000	2 semanas	\$ 50 000
M	2 semanas	1 semana	\$100 000	\$ 200 000	1 semana	\$100 000
N	6 semanas	3 semanas	\$330 000	\$ 510 000	3 semanas	\$ 60 000

■ **TABLA 9.10 Tabla inicial para analizar el costo del proyecto de Reliable**

Actividad de quiebre	Costo de quiebre	Longitud de la ruta					
		ABCDGHM	*ABCEHM*	*ABCEFJKN*	*ABCEFJLN*	*ABCIJKN*	*ABCIJLN*
		40	31	43	44	41	42

El problema: ¿Cuál es la manera menos costosa de acelerar algunas actividades para reducir la duración (estimada) del proyecto hasta el nivel especificado (40 semanas)?

Una técnica para resolver este problema es el **análisis de costo marginal**, que usa la última columna de la tabla 9.9 (junto con la tabla 9.8) para determinar la forma menos costosa de reducir la duración del proyecto una semana a la vez. El camino más sencillo para realizar este tipo de análisis es construir una tabla como la tabla 9.10 que enumere las rutas a través de la red del proyecto y la longitud actual de cada una. Para comenzar, esta información se puede copiar en forma directa de la tabla 9.8.

Como la cuarta trayectoria de la tabla 9.10 es la más larga (44 semanas), la única manera de reducir la duración del proyecto una semana es reducir la duración de las actividades en esta ruta una semana. Cuando se compara el costo de quiebre por semana ahorrada que se dio en la última columna de la tabla 9.9 para estas actividades, el costo menor es 30 000 dólares para la actividad *J*. (Observe que la actividad *I* con este mismo costo no está en esta ruta.) Por tanto, el primer cambio es acelerar la actividad *J* lo suficiente como para reducir su duración una semana.

Este cambio da como resultado una reducción de la longitud de cada trayectoria que incluye la actividad *J* —tercera, cuarta, quinta y sexta rutas de la tabla 9.10— una semana, como se muestra en el segundo renglón de la tabla 9.11. Como la cuarta ruta todavía es la más larga (43 semanas), se repite el proceso para encontrar la actividad en ella cuya aceleración cueste menos. De nuevo es la actividad *J*, puesto que la penúltima columna de la tabla 9.9 indica que se permite una reducción máxima de dos semanas de esta actividad. Esta segunda reducción de una semana de la actividad *J* conduce al tercer renglón de la tabla 9.11.

En este punto, la cuarta trayectoria todavía es la más larga (42 semanas), pero la actividad *J* no puede reducirse más. Entre las otras actividades de esta ruta, la *F* es ahora la de reducción menos costosa (40 000 dólares por semana), según la última columna de la tabla 9.9. Por tanto, se acelera esta actividad una semana para obtener el cuarto renglón de la tabla 9.11 y luego —debido a que se permite una reducción máxima de dos semanas— se acelera otra semana para obtener el último renglón de la tabla.

La trayectoria más larga —empate entre la primera, cuarta y sexta rutas— tiene ahora la longitud deseada de 40 semanas, por lo que se requiere más quiebre. (Si se necesitara, el siguiente paso requeriría estudiar las actividades en las tres trayectorias para encontrar la manera menos costosa de acortar las tres rutas una semana.) El costo total del quiebre de las actividades *J* y *F* para reducir la duración del proyecto a 40 semanas se calcula al sumar los costos de la segunda columna de la tabla 9.11, lo que da un total de 140 000 dólares. En la figura 9.30 se muestra la red de proyecto que resulta, donde las flechas más oscuras indican las rutas críticas.

■ **TABLA 9.11 Tabla final para realizar el análisis de costo marginal del proyecto de Reliable**

Actividad de quiebre	Costo de quiebre	Longitud de la ruta					
		ABCDGHM	*ABCEHM*	*ABCEFJKN*	*ABCEFJLN*	*ABCIJKN*	*ABCIJLN*
		40	31	43	44	41	42
J	$30 000	40	31	42	43	40	41
J	$30 000	40	31	41	42	39	40
F	$40 000	40	31	40	41	39	40
F	$40 000	40	31	39	40	39	40

En la figura 9.30 se muestra que reducir las duraciones de las actividades F y J a sus tiempos de quiebre genera *tres* rutas críticas a través de la red. La razón es que, como se encontró antes en el último renglón de la tabla 9.11, las tres trayectorias empatan como la más larga, cada una con longitud de 40 semanas.

Con redes más grandes, el análisis de costo marginal se puede convertir en algo difícil de manejar. Sería deseable contar con un procedimiento más eficiente para manejar proyectos grandes. Por esta razón, el procedimiento CPM estándar es aplicar *programación lineal*, por lo general mediante un paquete de software especializado que explote la estructura de este modelo de optimización de redes.

Uso de programación lineal para tomar decisiones de quiebre

El problema de encontrar la manera menos costosa de reducir actividades se puede establecer en una forma más familiar para programación lineal de la siguiente manera.

> **Enunciado del problema:** sea Z el costo total de reducir actividades. Entonces, el problema es minimizar Z sujeta a la restricción de que la duración del proyecto debe ser menor o igual que el tiempo deseado por el director del proyecto.

Las variables de decisión naturales son

x_j = reducción de la duración de la actividad j debido al quiebre de esta actividad, para $j = A, B \ldots, N$.

Al usar la última columna de la tabla 9.9, la función objetivo que se debe minimizar es

$$Z = 100\ 000x_A + 50\ 000x_B + \cdots + 60\ 000x_N.$$

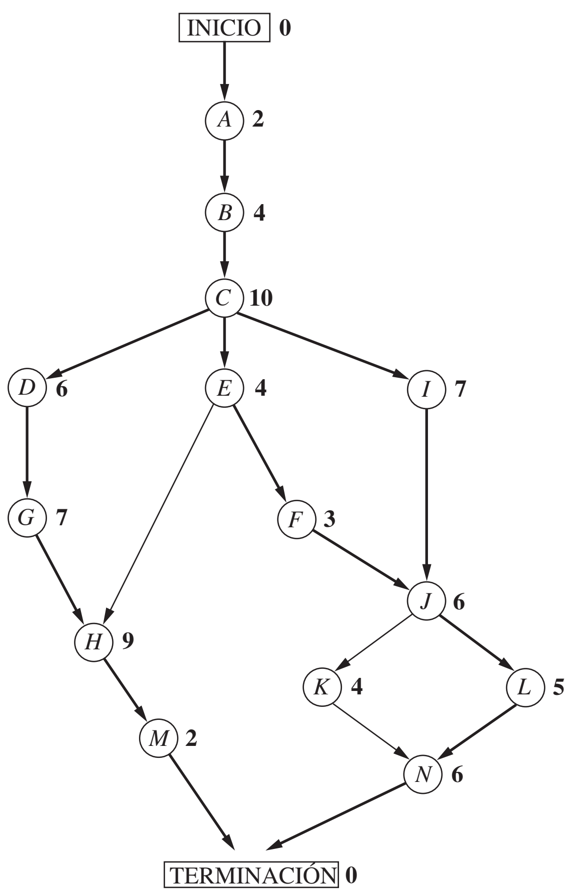


FIGURA 9.30
Red de proyecto si se hace un quiebre completo de las actividades J y F—con el resto de las actividades normales— del proyecto de Reliable. Las flechas más gruesas muestran las diferentes rutas críticas a través de la red.

Cada una de las 14 variables de decisión del lado derecho debe estar restringida a valores no negativos que no excedan el máximo incluido en la penúltima columna de la tabla 9.9.

Para imponer la restricción de que la duración del proyecto debe ser menor o igual que el valor deseado (40 semanas), sea

$y_{\text{TERMINACIÓN}}$ = duración del proyecto, es decir, el tiempo en el que se alcanza el nodo TERMINACIÓN de la red de proyecto.

La restricción es, entonces,

$$y_{\text{TERMINACIÓN}} \leq 40.$$

Para ayudar al modelo de programación lineal en la asignación del valor adecuado de $y_{\text{TERMINACIÓN}}$, dados los valores de $x_A, x_B, \ldots, x_N$, es conveniente introducir en el modelo las siguientes variables adicionales:

y_j = tiempo de inicio de la actividad j (para $j = B, C, \ldots, N$), dados los valores de $x_A, x_B, \ldots, x_N$.

(No es necesaria una variable de este tipo para la actividad A, puesto que una actividad que inicia el proyecto tiene el valor automático de 0.) Si el nodo TERMINACIÓN se maneja como otra actividad —con duración cero—, como se hará, esta definición de y_j para la actividad TERMINACIÓN también se ajusta a la definición de $y_{\text{TERMINACIÓN}}$ que se dio en el párrafo anterior.

El tiempo de inicio de cada actividad (incluso TERMINACIÓN) tiene relación directa con el tiempo de inicio y la duración de cada predecesor inmediato, como se resume a continuación.

Para cada actividad ($B, C, \ldots, N$, TERMINACIÓN) y cada predecesor inmediato,
Tiempo de inicio de esta actividad ≥ (tiempo de inicio + duración) de este predecesor inmediato.

Más aún, si se usan los tiempos normales de la tabla 9.9, la duración de cada actividad está dada por la siguiente fórmula:

Duración de la actividad j = su tiempo normal $- x_j$,

Para ilustrar estas relaciones, considere la actividad F en la red de proyecto (figuras 9.28 o 9.30).

Predecesor inmediato de la actividad F:
Actividad E, con duración $= 4 - x_E$.

Relaciones entre estas actividades:

$$y_F \geq y_E + 4 - x_E.$$

En consecuencia, la actividad F no puede comenzar hasta que empiece la actividad E y complete su duración de $4 - x_E$.

Ahora considere la actividad J, que tiene dos predecesores inmediatos.

Predecesores inmediatos de la actividad J:
Actividad F, con duración $= 5 - x_F$.
Actividad I, con duración $= 7 - x_I$.

Relaciones entre estas actividades:

$$y_J \geq y_F + 5 - x_F,$$
$$y_J \geq y_I + 7 - x_I.$$

Estas desigualdades juntas indican que la actividad j no puede comenzar hasta que ambos predecesores hayan terminado.

Al incluir estas relaciones para todas las actividades como restricciones, se obtiene el modelo completo de programación lineal que se presenta a continuación.

$$\text{Minimizar} \quad Z = 100\,000x_A + 50\,000x_B + \cdots + 60\,000x_N,$$

sujeta a las siguientes restricciones:

1. Restricciones de reducción máxima:
Si se usa la penúltima columna de la tabla 9.9,

$$x_A \leq 1, x_B \leq 2, \ldots, x_N \leq 3.$$

2. Restricciones de no negatividad:

$$x_A \geq 0, x_B \geq 0, \ldots, x_N \geq 0$$
$$y_B \geq 0, y_C \geq 0, \ldots, y_N \geq 0, y_{\text{TERMINACIÓN}} \geq 0.$$

3. Restricciones de tiempo de inicio:
Como se describió antes de la función objetivo, excepto por la actividad A (inicio al proyecto), existe una de estas restricciones para cada actividad con un solo predecesor inmediato (actividades B, C, D, E, F, G, I, K, L, M) y dos para cada actividad con dos predecesores inmediatos (actividades H, J, N, TERMINACIÓN), como se muestra en seguida.

Un predecesor inmediato	Dos predecesores inmediatos
$y_B \geq 0 + 2 - x_A$	$y_H \geq y_G + 7 - x_G$
$y_C \geq y_B + 4 - x_B$	$y_H \geq y_E + 4 - x_E$
$y_D \geq y_C + 10 - x_C$	$\vdots$
$\vdots$	$y_{\text{TERMINACIÓN}} \geq y_M + 2 - x_M$
$y_M \geq y_H + 9 - x_H$	$y_{\text{TERMINACIÓN}} \geq y_N + 6 - x_N$

(En general, el número de restricciones de tiempo de inicio es el número de predecesores inmediatos puesto que cada predecesor inmediato contribuye con una restricción de este tipo.)

4. Restricción de la duración del proyecto:

$$y_{\text{TERMINACIÓN}} \leq 40.$$

En la figura 9.31 se muestra la manera en que se puede formular este problema como un modelo de programación lineal en una hoja de cálculo. Las decisiones que deben tomarse se muestran en las celdas que cambian, TiempoDeInicio (I6:I19), ReduccionDeTiempo (J6:J19) y TiempoDeTerminacionDelProyecto (I22). Las columnas B a la H corresponden a las columnas de la tabla 9.9. Como lo indican las ecuaciones de la mitad inferior de la figura, las columnas G y H se calculan de manera directa. Las ecuaciones de la columna K expresan el hecho de que el tiempo de terminación de cada actividad es su tiempo de inicio *más* su tiempo normal *menos* su reducción debida al quiebre. La ecuación introducida en la celda objetivo CostoTotal (I24) suma todos los costos normales más los costos adicionales por el quiebre para obtener el costo total.

El último conjunto de restricciones del cuadro de diálogo de Solver, ReduccionDeTiempo (J6:J19) ≤ ReduccionDeTiempoMax (G6:G19) especifica que la reducción de tiempo de cada actividad no puede exceder el máximo que se presentó en la columna G. Las dos restricciones anteriores TiempoDeTerminacionDelProyecto (I22) ≥ MTerminacion (K18) y TiempoDeTerminacionDelProyecto (I22) ≥ NTERMINACION (K19) indican que el proyecto no puede terminarse hasta que cada uno de los dos predecesores inmediatos (actividades M y N) terminen. La restricción TiempoDeTerminacionDelProyecto (I22) ≤ TiempoMax (K22) es importante y especifica que el proyecto debe terminar en 40 semanas.

Todas las restricciones que involucran TiempoDeInicio (I6:I19) son *restricciones de tiempo de inicio* que especifican que una actividad no puede comenzar hasta que sus predecesores inmediatos terminen. Por ejemplo, la primera restricción BInicio (I7) ≥ ATerminacion (K6) indica que la actividad B no puede comenzar hasta que la A (su predecesor inmediato) termine. Cuando una actividad tiene más de un predecesor inmediato, hay una restricción para cada uno de ellos. Por ejemplo, la actividad H tiene como predecesores inmediatos a E y G. En consecuencia, H tiene dos restricciones de tiempo de inicio, HInicio (I13) ≥ ETerminacion (K10) y HInicio (I13) ≥ GTerminacion (K12).

Habrá observado que la forma ≥ de las *restricciones de tiempo de inicio* permite un retraso en el inicio de una actividad una vez que sus predecesores inmediatos han terminado. Aunque esta demora es factible en el modelo, no puede ser óptima para las actividades de la ruta crítica, puesto

que puede aumentar el costo total (seguro necesitarán un quiebre adicional para cumplir con la restricción de duración del proyecto). Por tanto, una solución óptima del modelo no tendrá estas demoras, excepto quizás en el caso de actividades fuera de la ruta crítica.

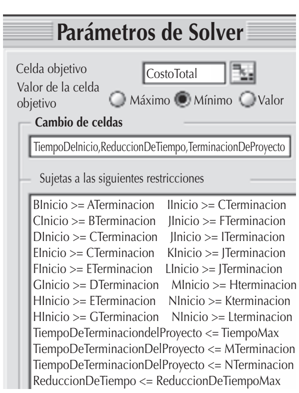
Las columnas I y J de la figura 9.31 muestran la solución óptima que se obtiene después de hacer clic en el botón de Solver. (Observe que esta solución incluye un retraso —la actividad *K* comienza en 30 aunque su único predecesor inmediato, la actividad *J*, termina en 29— pero ello no importa pues *K* no está en la ruta crítica.) Esta solución corresponde a la que se despliega en la figura 9.30 que se obtuvo con el análisis de costos marginales.

Si desea ver **otro ejemplo** que ilustre tanto el enfoque de análisis de costo marginal como el de programación lineal para aplicar el método CPM de trueques entre tiempo y costo, en la sección Worked Examples en el sitio en internet de este libro se proporciona uno.

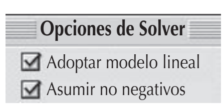
FIGURA 9.31
La hoja de cálculo muestra la aplicación del método CPM de trueques entre tiempo y costo en el proyecto de Reliable, donde las columnas *I* y *J* muestran la solución óptima que se obtuvo al usar Excel Solver con los elementos que se presentan en el cuadro de diálogo de Solver.

	A	B	C	D	E	F	G	H	I	J	K
1		Problema de programación del proyecto de la Reliable Construction Co. con trueques entre tiempo y costo									
2								Costo			
3							Máxima	de quiebre			
4			Tiempo		Costo		reducción	por semana	Tiempo de	Reducción	Tiempo de
5		Actividad	Normal	Quiebre	Normal	Quiebre	de tiempo	ahorrado	inicio	de tiempo	terminación
6		A	2	1	$180 000	$280 000	1	$100 000	0	0	2
7		B	4	2	$320 000	$420 000	2	$50 000	2	0	6
8		C	10	7	$620 000	$860 000	3	$80 000	6	0	16
9		D	6	4	$260 000	$340 000	2	$40 000	16	0	22
10		E	4	3	$410 000	$570 000	1	$160 000	16	0	20
11		F	5	3	$180 000	$260 000	2	$40 000	20	2	23
12		G	7	4	$900 000	$1 020 000	3	$40 000	22	0	29
13		H	9	6	$200 000	$380 000	3	$60 000	29	0	38
14		I	7	5	$210 000	$270 000	2	$30 000	16	0	23
15		J	8	6	$430 000	$490 000	2	$30 000	23	2	29
16		K	4	3	$160 000	$200 000	1	$40 000	30	0	34
17		L	5	3	$250 000	$350 000	2	$50 000	29	0	34
18		M	2	1	$100 000	$200 000	1	$100 000	38	0	40
19		N	6	3	$330 000	$510 000	3	$60 000	34	0	40
20											
21											Tiempo máx
22							Terminación del proyecto		40	<=	40
23											
24								Costo total	$4 690 000		

Nombre de rango	Celdas
ATerminación	K6
AInicio	I6
BTerminación	K7
BInicio	I7
CTerminación	K8
CostoDeQuiebre	F6:F19
CostoDeQuiebre-AhorradoPorSemana	H6:H19
TiempoDeQuiebre	D6:D19
CInicio	I8
DTerminación	K9
DInicio	I9
Eterminación	K10
EInicio	I10
FTerminación	K11
TiempoDeTerminación	K6:K19
FInicio	I11
GTerminación	K12
GInicio	I12
HTerminación	K13
HInicio	I13
ITerminación	K14
IInicio	I14
JTerminación	K15
JInicio	I15
KTerminación	K16
KInicio	I16
LTerminación	K17
LInicio	I17
TiempoMax	K22
ReducciónDeTiempoMáx	G6:G19
MTerminación	K18
MInicio	I18
NTerminación	K19
CostoNormal	E6:E19
TiempoNormal	C6:C19
NInicio	I19
TiempoDeTerminación DelProyecto	I22
TiempoDeInicio	I6:I19
ReduccionDeTiempo	J6:J19
CostoTotal	I24



	G	H
3	Reducción	Costo de quiebre
4	de tiempo	por semana
5	máxima	ahorrado
6	=TiempoNormal-TiempoDeQuiebre	=(CostoDeQuiebre-CostoNormal)/ReduccionDeTiempoMax
7	=TiempoNormal-TiempoDeQuiebre	=(CostoDeQuiebre-CostoNormal)/ReduccionDeTiempoMax
8	=TiempoNormal-TiempoDeQuiebre	=(CostoDeQuiebre-CostoNormal)/ReduccionDeTiempoMax
9	=TiempoNormal-TiempoDeQuiebre	=(CostoDeQuiebre-CostoNormal)/ReduccionDeTiempoMax
10	:	:
11	:	:



	K
4	Tiempo
5	de terminación
6	=TiempoDeInicio+TiempoNormal-ReduccionDeTiempo
7	=TiempoDeInicio+TiempoNormal-ReduccionDeTiempo
8	=TiempoDeInicio+TiempoNormal-ReduccionDeTiempo
9	=TiempoDeInicio+TiempoNormal-ReduccionDeTiempo
10	:
11	:

	H	I
24	CostoTotal	=SUMA(CostoNormal)+SUMAPRODUCTO(CostoDeQuiebreAhorradoPorSemana,ReduccionDeTiempo)

9.9 CONCLUSIONES

Las redes de cierto tipo surgen en una amplia variedad de contextos. Las representaciones de redes son muy útiles para visualizar las relaciones y conexiones entre las componentes del sistema. Con frecuencia debe enviarse un flujo de algún tipo a través de la red y es necesario tomar una decisión sobre la mejor manera de hacerlo. En este capítulo se introdujeron dos tipos de modelos y algoritmos de optimización de redes que constituyen una herramienta poderosa para tomar esas decisiones.

El problema del flujo de costo mínimo juega un papel central entre estos nuevos modelos de optimización de redes, tanto por su extensa aplicación como porque se puede resolver con gran eficiencia por el método símplex de redes. Dos de sus casos especiales, incluidos en este capítulo —el problema de la ruta más corta y el del flujo máximo—, también son modelos importantes de optimización de redes, al igual que los casos especiales que se presentaron en el capítulo 8 (el problema de transporte y el problema de asignación).

Al mismo tiempo que todos estos modelos se ocupan de optimizar la *operación* de una red *existente*, el problema del árbol de mínima expansión es un ejemplo sobresaliente de un modelo para optimizar el *diseño* de una *nueva* red.

El método CPM de trueques entre tiempo y costo proporciona una forma poderosa de utilizar un modelo de optimización de redes para diseñar un proyecto de manera que pueda cumplir con su fecha límite con un costo total mínimo.

En este capítulo apenas se ha tocado la superficie de lo que hasta la fecha se ha desarrollado en el campo de la metodología de redes. Por su naturaleza combinatoria, con frecuencia los problemas de redes son difíciles de resolver, pero se han logrado grandes avances en el desarrollo de técnicas poderosas de modelado y de metodologías de solución que incluso amplían el panorama de nuevas e importantes aplicaciones. En realidad, los nuevos algoritmos han permitido resolver con éxito algunos problemas complejos de redes de gran tamaño.

REFERENCIAS SELECCIONADAS

1. Ahuja, R. K., T. L. Magnanti y J. B. Orlin: *Network Flows: Theory, Algorithms, and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
2. Bertsekas, D. P.: *Network Optimization: Continuous and Discrete Models*, Athena Scientific Publishing, Belmont, MA, 1998.
3. Cai, X. y C. K. Wong: *Time Varying Network Optimization*, Springer, Nueva York, 2007.
4. Dantzig, G. B. y M. N. Thapa: *Linear Programming 1: Introduction*, Springer, Nueva York, 1997, cap. 9.
5. Hillier, F. S. y M. S. Hillier: *Introduction to Management Science: A Modeling and Case Studies Approach with Spreadsheets*, 3a. ed., McGraw-Hill/Irwin, Burr Ridge, IL, 2008, cap. 6.
6. Magnanti, T. L. y R. T. Wong: “Network Design and Transportation Planning: Models and Algorithms”, *Transportation Science*, **18:** 1-55, 1984.
7. Vanderbei, R. J.: *Linear Programming*: *Foundations and Extensions*, 3a. ed., Springer, Nueva York, 2008, caps. 14 y 15.

Algunas aplicaciones ganadoras de premios de los modelos de optimización de redes:

(El vínculo a estos artículos se encuentra en nuestro sitio web, en internet www.mhhe.com/hillier) se proporciona una liga a los artículos siguientes.

A1. Ben-Khedher, N., J. Kintanar, C. Queille y W. Stripling: “Schedule Optimization at SNCF: From Conception to Day of Departure”, en *Interfaces*, 28(1): 6-23, enero-febrero de 1998.

A2. Blais, J. Y., J. Lamont y J. M. Rousseau: “The HASTUS Vehicle and Manpower Scheduling at the Société de transport de la Communauté urbaine de Montréal”, en *Interfaces*, **20**(1): 26-42, enero-febrero de 1990.

A3. Cosares, S., D. N. Deutsch, I. Saniee y O. J. Wasem: “SONET Toolkit: A Decision-Support System for Designing Robust and Cost-Effective Fiber-Optic Networks”, *Interfaces*, **25**(1): 20-40, enero-febrero de 1995.

A4. Huisingh, J. L., H. M. Yamauchi y R. Zimmerman, "Saving Federal Tax Dollars", en *Interfaces*, **31**(5): 13-23, septiembre-octubre de 2001.
A5. Klingman, D., N. Phillips, D. Steiger y W. Young: "The Successful Deployment of Management Science throughout Citgo Petroleum Corporation", en *Interfaces*, **17**(1): 4-25, enero-febrero de 1987.
A6. Powell, W. B., Y. Sheffi, K. S. Nickerson, K. Butterbaugh y S. Atherton: "Maximizing Profits for North American Van Lines' Truckload Division: A New Framework for Pricing and Operations", en *Interfaces*, **18**(1): 21-41, enero-febrero de 1988.
A7. Prior, R. C., R. L. Slavens, J. Trimarco, V. Akgun, E. G. Feitzinger y C. F. Hong: "Menlo Worldwide Forwarding Optimizes Its Network Routing", en *Interfaces*, **34**(1): 26-38, enero-febrero de 2004.
A8. Srinivasan, M. M., W. D. Best y S. Chandrasekaran: "Warner Robins Air Logistics Center Streamlines Aircraft Repair and Overhaul", en *Interfaces*, **37**(1): 7-21, enero-febrero de 2007.
A9. Vasquez-Marquez, A.: "American Airlines Arrival Slot Allocation System (ASAS)", en *Interfaces*, **21**(1): 42-61, enero-febrero de 1991.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN NUESTRO SITIO EN INTERNET (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos para el capítulo 9

Un ejemplo de demostración en el OR Tutor:

Método símplex para redes (*Network Simplex Method*)

Una rutina interactiva en el IOR Tutorial:

Método símplex para redes (*Network Simplex Method*), Interactivo

Complemento de Excel:

Premium Solver for Education

Archivos de "Ch. 9—Network Opt Models" para resolver los ejemplos:

Archivos de Excel
Archivo de LINGO/LINDO
Archivo de MPL/CPLEX

Glosario del capítulo 9

Vea en el apéndice 1 la documentación del software.

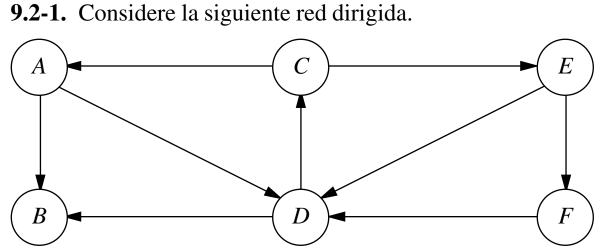
PROBLEMAS

Los símbolos a la izquierda de algunos problemas (o de sus incisos) significan lo siguiente:

D: El ejemplo de demostración indicado en las Ayudas de Aprendizaje puede ser útil.
I: Se sugiere al lector que utilice el procedimiento interactivo que se acaba de indicar (la impresión registra su trabajo).
C: Use la computadora con cualquier software disponible (o según indique el instructor) para resolver el problema.

Un asterisco en el número del problema indica que al final del libro se proporciona, al menos, una respuesta parcial.

9.2-1. Considere la siguiente red dirigida.



***a*)** Encuentre una trayectoria dirigida del nodo A al nodo F y después identifique otras tres trayectorias no dirigidas del nodo A al nodo F.

b) Encuentre tres ciclos dirigidos. Después identifique un ciclo no dirigido que incluya todos los nodos.

c) Identifique un conjunto de arcos que formen un árbol de expansión.

d) Use el proceso que se presenta en la figura 9.3 para construir un árbol con un arco a la vez hasta tener un árbol de expansión. Después repita el proceso para obtener otro árbol de expansión. [No duplique el árbol de expansión identificado en el inciso *c*).]

9.3-1. Lea el artículo de referencia que describe a detalle el estudio de IO que se resume en el recuadro de aplicación que se presentó en la sección 9.3. Describa brevemente qué modelo de optimización de redes se aplicó en ese estudio. Enliste los diferentes beneficios financieros y de otro tipo que resultaron de ese estudio.

9.3-2. Usted debe hacer un viaje en automóvil a una ciudad que nunca ha visitado. Estudia un plano para determinar la ruta más corta hasta su destino. Según la ruta que elija, hay otras cinco ciudades (llamadas A, B, C, D, E) por las que puede pasar en el camino. El plano muestra las millas de cada carretera que son una conexión directa entre dos ciudades sin que otra intervenga. Estas cifras se resumen en la siguiente tabla, donde un guión indica que no hay conexión directa sin pasar por otras ciudades.

	Millas entre ciudades adyacentes					
Ciudad	**A**	**B**	**C**	**D**	**E**	**Destino**
Origen	40	60	50	—	—	—
A		10	—	70	—	—
B			20	55	40	—
C				—	50	—
D					10	60
E						80

a) Formule éste como un problema de la ruta más corta al trazar una red donde los nodos son ciudades, los arcos son carreteras y los números la distancia en millas.

b) Use el algoritmo descrito en la sección 9.3 para resolver este problema de la ruta más corta.

C ***c***) Formule y resuelva un modelo en hoja de cálculo para este problema.

d) Si cada número de la tabla representa su costo (en dólares) de manejar de una ciudad a la siguiente, ¿obtiene la ruta de costo mínimo con la respuesta del inciso *b*) o *c*)?

e) Si cada número de la tabla representa su *tiempo* (en minutos) para manejar de una ciudad a la siguiente, ¿obtiene la ruta de tiempo mínimo con la respuesta del inciso *b*) o *c*)?

9.3-3. Una compañía aérea local piensa comprar un tractor nuevo para mover el tren de carros que llevan y traen el equipaje de los aviones que aterrizan en un pequeño aeropuerto que está en pleno crecimiento. Dentro de tres años se instalará un nuevo sistema mecanizado de transporte de equipaje, por lo que después no se necesitará el tractor. No obstante, tendrá una carga de trabajo pesada y los costos de operación y mantenimiento aumentarán rápido y podría resultar costeable reemplazarlo en uno o dos años. La siguiente tabla proporciona los costos descontados netos totales asociados con la compra del tractor (precio de compra menos valor de venta del tractor en uso más costos de operación y mantenimiento) al final del año *i* y si se reemplaza al final del año *j* (donde el momento presente es el año 0).

		j		
		1	**2**	**3**
	0	$13 000	$28 000	$48 000
i	1		$17 000	$33 000
	2			$20 000

El problema es determinar en qué momento (si existe) debe reemplazarse el tractor para minimizar el costo total durante los tres años.

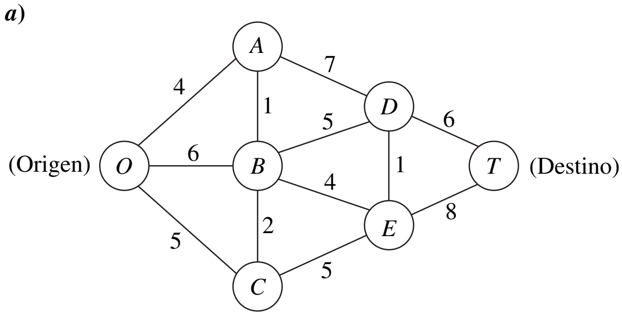
a) Formule éste como un problema de ruta más corta.

b) Utilice el algoritmo descrito en la sección 9.3 para resolver este problema de ruta más corta.

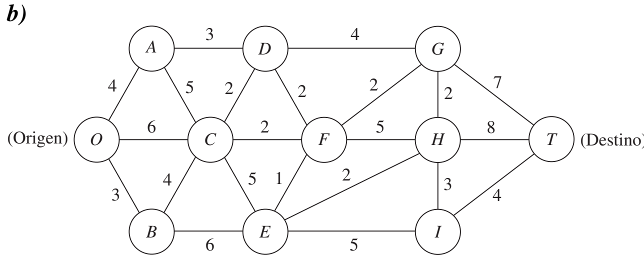
C ***c***) Formule y resuelva el modelo en hoja de cálculo.

9.3-4.* Utilice el algoritmo descrito en la sección 9.3 para encontrar la *ruta más corta* a través de las redes *a*) y *b*), *en las cuales* los números representan las distancias reales entre los nodos correspondientes.

a)

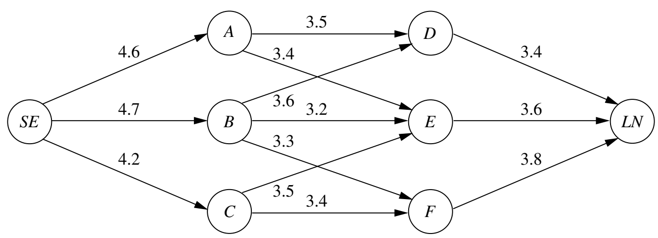


b)



9.3-5. Formule el problema de la ruta más corta como uno de programación lineal.

9.3-6. Un vuelo de Speedy Airlines está a punto de despegar de Seattle sin escalas a Londres. Existe cierta flexibilidad para elegir la ruta precisa, según las condiciones del clima. La siguiente red describe las rutas posibles consideradas, donde SE y LN son Seattle y Londres, respectivamente, y los otros nodos representan varios lugares intermedios.



El viento a lo largo de cada arco afecta de manera considerable el tiempo de vuelo, y, por ende, el consumo de combustible. Con base en el informe meteorológico actual, junto a los arcos se muestran los tiempos de vuelo (en horas). Debido al alto costo del combustible, la administración ha adoptado la política de elegir la ruta que minimiza el tiempo total de vuelo.

a) ¿Qué juega el papel de "distancias" en la interpretación de este problema?
b) Use el algoritmo descrito en la sección 9.3 para resolver este problema de la ruta más corta.
C ***c***) Formule y resuelva el modelo en una hoja de cálculo.

9.3-7. La compañía Quick ha averiguado que un competidor planea lanzar un nuevo tipo de producto con ventas potenciales muy grandes. Quick ha trabajado en un producto similar programado para salir dentro de 20 meses. Sin embargo, la investigación está casi terminada y ahora la administración quiere lanzar el producto más rápidamente para hacer frente a la competencia.

Se deben superar cuatro etapas independientes que incluyen lo que falta de la investigación que por el momento se lleva a cabo a paso normal. No obstante, cada etapa se puede realizar en un nivel de prioridad o de concentración para acelerar la terminación y éstos son los únicos niveles considerados en las últimas tres etapas. Los tiempos que se requieren para cada nivel se muestran en la siguiente tabla. (Los tiempos entre paréntesis correspondientes al nivel normal se han eliminado por ser muy largos.)

	Tiempo			
Nivel	**Investigación restante**	**Desarrollo**	**Diseño del sistema de manufactura**	**Inicio de producción y distribución**
Normal	5 meses	(4 meses)	(7 meses)	(4 meses)
Prioridad	4 meses	3 meses	5 meses	2 meses
Quiebre	2 meses	2 meses	3 meses	1 mes

La administración ha destinado 50 millones de dólares para las cuatro etapas. El costo (en millones de dólares) de cada fase en los diferentes niveles bajo consideración es:

	Costo			
Nivel	**Investigación restante**	**Desarrollo**	**Diseño del sistema de manufactura**	**Inicio de producción y distribución**
Normal	\$5 millones	—	—	—
Prioridad	\$9 millones	\$10 millones	\$14 millones	\$6 millones
Quiebre	\$14 millones	\$15 millones	\$19 millones	\$9 millones

La administración desea determinar el nivel al que debe realizar cada una de las cuatro etapas para minimizar el tiempo total hasta la comercialización del producto sujeto a las restricciones de presupuesto.

a) Formule éste como un problema de la ruta más corta.
b) Utilice el algoritmo descrito en la sección 9.3 para resolverlo.

9.4-1.[*] Reconsidere las redes que se muestran en el problema 9.3-4. Utilice el algoritmo descrito en la sección 9.4 para encontrar el *árbol de expansión mínima* de cada una de ellas.

9.4-2. La maderera Wirehouse talará árboles en ocho zonas de la misma área. Pero antes debe desarrollar un sistema de caminos de tierra para tener acceso a cualquier zona desde cualquier otra. La distancia (en millas) entre cada par de zonas es:

		Distancia entre pares de zonas							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	1	—	1.3	2.1	0.9	0.7	1.8	2.0	1.5
	2	1.3	—	0.9	1.8	1.2	2.6	2.3	1.1
	3	2.1	0.9	—	2.6	1.7	2.5	1.9	1.0
Zona	4	0.9	1.8	2.6	—	0.7	1.6	1.5	0.9
	5	0.7	1.2	1.7	0.7	—	0.9	1.1	0.8
	6	1.8	2.6	2.5	1.6	0.9	—	0.6	1.0
	7	2.0	2.3	1.9	1.5	1.1	0.6	—	0.5
	8	1.5	1.1	1.0	0.9	0.8	1.0	0.5	—

El problema es determinar los pares de zonas entre los que deben construirse caminos para conectar todas con una longitud de caminos total mínima.

a) Describa cómo se ajusta este problema a la descripción del problema del árbol de expansión mínima.
b) Utilice el algoritmo descrito en la sección 9.4 para resolverlo.

9.4-3. El Premiere Bank ha decidido conectar terminales de computadora de cada sucursal a la computadora central de su oficina matriz mediante líneas telefónicas especiales con dispositivos de telecomunicaciones. No es necesario que la línea telefónica de una sucursal esté conectada directamente con la oficina matriz. La conexión puede ser indirecta a través de otra sucursal que esté conectada (directa o indirectamente) a la matriz. El único requisito es que exista alguna ruta que conecte a todas las sucursales con la oficina matriz.

El cargo por las líneas telefónicas especiales es directamente proporcional a la distancia cableada, donde la distancia (en millas) entre cada par de oficinas es:

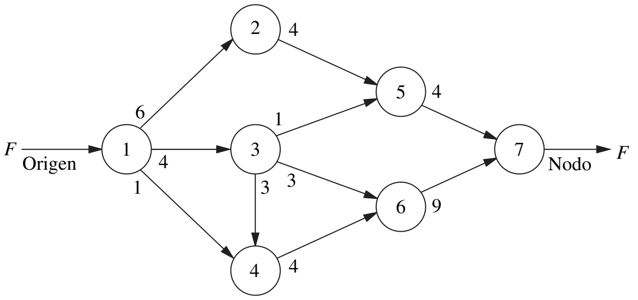
	Distancia entre pares de oficinas					
	Principal	**S.1**	**S.2**	**S.3**	**S.4**	**S.5**
Oficina principal	—	190	70	115	270	160
Sucursal 1	190	—	100	110	215	50
Sucursal 2	70	100	—	140	120	220
Sucursal 3	115	110	140	—	175	80
Sucursal 4	270	215	120	175	—	310
Sucursal 5	160	50	220	80	310	—

La administración desea determinar qué pares de sucursales conectar directamente con las líneas telefónicas especiales para que todas queden conectadas (de modo directo o indirecto) a la oficina matriz con un costo total mínimo.

a) Explique cómo se ajusta este problema a la descripción del problema del árbol de expansión mínima.
b) Utilice el algoritmo descrito en la sección 9.4 para resolver este problema.

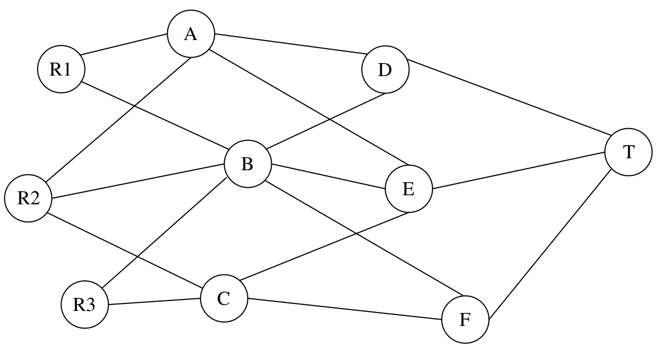
9.5-1.[*] Para la red mostrada a continuación, utilice el algoritmo de la trayectoria de aumento descrita en la sección 9.5 para encontrar el patrón de flujo que proporciona el *flujo máximo* del nodo origen al nodo destino, dado que la capacidad a través del arco que va del nodo

i al nodo *j* es el número más cercano al nodo *i* del arco entre estos nodos. Muestre su trabajo.



9.5-2. Formule el problema de flujo máximo como un problema de programación lineal.

9.5-3. El siguiente diagrama describe un sistema de acueductos que se origina en tres ríos (Rl, R2 y R3) y termina en una ciudad importante (nodo T), donde los otros nodos son puntos de unión del sistema.



Utilice unidades de miles de acres-pie; las siguientes tablas muestran la cantidad máxima de agua que puede bombearse, a través de cada acueducto, cada día.

De \ A	A	B	C
R1	130	115	—
R2	70	90	110
R3	—	140	120

De \ A	D	E	F
A	110	85	—
B	130	95	85
C	—	130	160

De \ A	T
D	220
E	330
F	240

La comisión del agua desea determinar el plan que maximice el flujo de agua hacia la ciudad.

a) Formule este problema como un problema de flujo máximo; identifique un origen, un destino y los nodos de trasbordo, y trace la red completa que muestre la capacidad de cada arco.

b) Use el algoritmo de la trayectoria de aumento que se presentó en la sección 9.5 para resolver este problema.

C ***c***) Formule y resuelva el modelo en una hoja de cálculo.

9.5-4. La Texago Corporation tiene cuatro campos de petróleo, cuatro refinerías y cuatro centros de distribución. Una fuerte huelga en la industria del transporte ha reducido de manera considerable la capacidad de Texago para enviar petróleo de sus campos a las refinerías y los productos derivados a los centros de distribución. Use unidades en miles de barriles de petróleo crudo (y su equivalente en productos refinados); las tablas siguientes muestran el número máximo de unidades que puede enviar al día de cada campo a cada refinería y de éstas a cada centro de distribución.

	Refinería			
Campo	**N. Orleans**	**Charleston**	**Seattle**	**San Luis**
Texas	11	7	2	8
California	5	4	8	7
Alaska	7	3	12	6
Medio oeste	8	9	4	15

	Centro de distribución			
Refinería	**Pittsburgh**	**Atlanta**	**Kansas City**	**San Francisco**
N. Orleans	5	9	6	4
Charleston	8	7	9	5
Seattle	4	6	7	8
San Luis	12	11	9	7

La administración de Texago desea elaborar un plan para determinar cuántas unidades debe enviar de cada campo petrolero a cada refinería y de cada refinería a cada centro de distribución de manera que se maximice el número total de unidades que llegan a los centros de distribución.

a) Bosqueje un plano que muestre la ubicación de los campos, refinerías y centros de distribución de Texago. Agregue el flujo del petróleo crudo y de los productos del petróleo a través de la red de distribución.

b) Dibuje de nuevo la red alineando en una columna los nodos de los campos, en otra los de refinerías y en una tercera los de centros de distribución. Después agregue arcos para mostrar el flujo posible.

c) Modifique la red del inciso *b*) para formular este problema como uno de flujo máximo con sólo una fuente, un destino y una capacidad de cada arco.

d) Utilice el algoritmo de la trayectoria de aumento de la sección 9.5 para resolver el problema de flujo máximo.

C ***e***) Formule y resuelva el modelo en una hoja de cálculo.

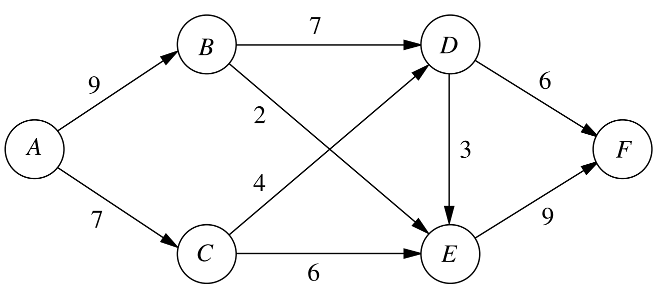
9.5-5. Una vía del sistema de ferrocarril Eura Railroad va de Faireparc, la ciudad industrial más importante, a Portstown, la ciudad portuaria más grande. Esta vía es utilizada de manera frecuente tanto por expresos de pasajeros como por trenes de carga. Los expresos se programan con detalle y tienen prioridad sobre los de carga más lentos (es un ferrocarril europeo), de manera que el tren de carga debe salir a una vía lateral cuando esté programado que uno de pasajeros lo rebase. Es necesario aumentar el servicio de carga, por lo que el problema es programar los trenes de forma que se maximice el número que se puede enviar cada día sin interferir con los horarios fijos de los expresos.

Los trenes de carga consecutivos deben mantener una diferencia de horarios de al menos 0.1 hora; se usa esta unidad de tiempo para programarlos, por lo cual el itinerario del día indica la posición relativa de cada tren de carga en los tiempos 0.0, 0.1, 0.2, . . . , 23.9. Se cuenta con S vías laterales entre Faireparc y Portstown, las cuales tienen el largo suficiente para recibir n_i trenes de carga ($i = 1, \ldots, S$). Se requiere t_i unidades de tiempo (redondeadas a

un entero) para que un tren de carga viaje de la vía lateral i al $i + 1$ —con t_0 = tiempo requerido para ir de Faireparc a la vía 1 y t_S es el tiempo de la vía lateral S a Portstown—. Se permite que un tren de carga pase o deje la vía lateral $i(i = 0, 1, \ldots, S)$ en el tiempo j ($j = 0.0, 0.1, \ldots, 23.9$) si, según lo programado, no lo alcanzará un expreso antes de llegar a la vía lateral $i + 1$ (sea $\delta_{ij} = 1$ si lo alcanzará y $\delta_{ij} = 0$ si no). También se requiere que un tren de carga espere si no hay lugar en las vías laterales siguientes, antes de que lo alcance un expreso.

Formule este problema como uno de flujo máximo e identifique cada nodo (sin incluir origen o destino) y cada arco con la capacidad de flujo para la representación de redes del problema. (*Sugerencia*: Utilice un conjunto diferente de nodos para cada uno de los 240 tiempos.)

9.5-6. Considere el problema de flujo máximo que se muestra en la siguiente red, en donde A es el nodo origen y F el nodo de demanda, mientras que las capacidades son los números que se muestran junto a los arcos dirigidos.



a) Utilice el algoritmo de la trayectoria de aumento descrito en la sección 9.5 para resolver este problema.

C ***b***) Formule y resuelva un modelo en hoja de cálculo para este problema.

9.6-1. Lea el artículo en el que se describe a detalle el estudio de IO que se resume en el recuadro de aplicación de la sección 9.6. Brevemente describa cómo el modelo de flujo de costo mínimo puede aplicarse a este estudio. Describa de manera breve cómo se aplicaron los modelos de optimización de red a este estudio. Posteriormente enliste los diferentes beneficios financieros y de otro tipo que arrojó dicho estudio.

9.6-2. Reconsidere el problema de flujo máximo que se presentó en el problema 9.5-6. Formule este problema como uno de flujo de costo mínimo, y agregue el arco $A \to F$. Use $\overline{F} = 20$.

9.6-3. Una compañía fabricará el mismo producto nuevo en dos plantas y después lo enviará a dos almacenes. La fábrica 1 puede enviar una cantidad ilimitada por ferrocarril sólo al almacén 1, mientras que la fábrica 2 puede mandar una cantidad ilimitada por la misma vía sólo al almacén 2. Sin embargo, se puede usar camiones de carga independientes para enviar hasta 50 unidades de cada fábrica a un centro de distribución desde el que se pueden enviar hasta 50 unidades a cada almacén. En la siguiente tabla se muestra el costo unitario de embarque de cada alternativa junto con las cantidades que se producirán en las fábricas y las cantidades que se necesitan en los almacenes.

A / De	Costo unitario de embarque			Producción
	Centro de distribución	Almacén 1	Almacén 2	
Fábrica 1	3	7	—	80
Fábrica 2	4	—	9	70
Centro de distribución		2	4	
Asignación		60	90	

a) Formule la representación de redes de este problema como un problema de flujo de costo mínimo.

b) Formule un modelo de programación lineal para este problema.

9.6-4. Reconsidere el problema 9.3-3. Ahora formule este problema como un problema de flujo de costo mínimo y muestre su representación de redes apropiada.

9.6-5. Makonsel es una compañía integral que produce bienes y los vende en sus propias tiendas. Después de producidos los bienes se colocan en dos almacenes hasta que las tiendas los necesitan. Se usan camiones para transportar los bienes a los almacenes y luego a las tres tiendas.

Utilice una carga completa de camión como unidad; la siguiente tabla muestra la producción mensual de cada planta, su costo de transporte por carga enviada a cada almacén y la cantidad máxima que se puede enviar al mes a cada uno.

A / De	Costo unitario de envío		Capacidad de envío		Producción
	Almacén 1	Almacén 2	Almacén 1	Almacén 2	
Planta 1	\$1 175	\$1 580	375	450	600
Planta 2	\$1 430	\$1 700	525	600	900

La siguiente tabla contiene la demanda mensual de cada tienda (T), el costo de transporte por camión desde cada almacén y la cantidad máxima que se puede enviar al mes desde cada uno.

A / De	Costo unitario de envío			Capacidad de envío		
	T1	T2	T3	T1	T2	T3
Almacén 1	\$1 370	\$1 505	\$1 490	300	450	300
Almacén 2	\$1 190	\$1 210	\$1 240	375	450	225
Demanda	450	600	450	450	600	450

La administración desea determinar un plan de distribución —número de cargas enviadas al mes de cada planta a cada almacén y de cada uno de éstos a cada tienda— de modo que se minimice el costo total de transporte.

a) Trace una red que describa la red de distribución de la compañía. Identifique en ella los nodos fuente, de trasbordo y de demanda.
b) Formule este problema como uno de flujo de costo mínimo y coloque todos los datos necesarios.
C ***c***) Formule y resuelva un modelo en hoja de cálculo.
C ***d***) Use una computadora para resolver este problema sin usar Excel.

9.6-6. La compañía Audiofile produce aparatos de sonido portátiles. Sin embargo, la administración ha decidido subcontratar la producción de las bocinas necesarias para dichos aparatos de sonido. Existen tres proveedores. Sus precios por cada embarque de 1 000 bocinas se muestran en la siguiente tabla.

Proveedor	Precio
1	$22 500
2	$22 700
3	$22 300

Además, cada proveedor cobrará un costo de envío. Cada embarque llegará a uno de los dos almacenes de la compañía. Cada proveedor tiene su propia fórmula para calcular este costo según las millas recorridas hasta el almacén. Estas fórmulas y los datos de las millas se muestran a continuación.

Proveedor	Cargo por envío
1	$300 + 40¢/milla
2	$200 + 50¢/milla
3	$500 + 20¢/milla

Proveedor	Almacén 1	Almacén 2
1	1 600 millas	400 millas
2	500 millas	600 millas
3	2 000 millas	1 000 millas

Cuando una de las dos fábricas requiere un embarque de bocinas para amenizar los bailes, contrata un camión para traerlo de los almacenes. El costo por embarque se presenta en la siguiente columna, junto con el número de embarques por mes que requiere cada planta.

	Costo unitario por envío	
	Fábrica 1	**Fábrica 2**
Almacén 1	$200	$700
Almacén 2	$400	$500
Demanda mensual	10	6

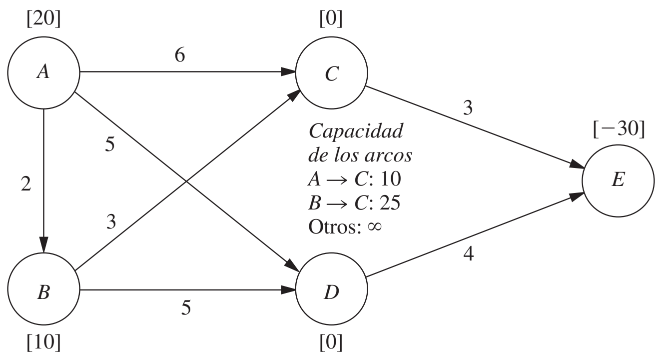
Cada proveedor puede surtir hasta 10 embarques por mes; pero debido a las limitaciones de transporte, cada uno puede enviar un máximo de sólo 6 embarques por mes a cada almacén. De manera similar, cada almacén puede enviar hasta 6 embarques por mes a cada fábrica.

Ahora, la administración desea desarrollar un plan mensual para determinar cuántos embarques (si son necesarios) ordenar a cada proveedor, cuántos de ellos deben ir a cada almacén y cuántos embarques debe enviar cada almacén a cada fábrica. El objetivo es minimizar la suma de los costos de compra (incluyendo los cargos de envío) y los costos de envío desde los almacenes a las fábricas.

a) Trace una red que describa la red de proveedores de Audiofile. Identifique en ella los nodos de suministro, trasbordo y demanda.
b) Formule este problema como uno de flujo de costo mínimo, con todos los datos necesarios en la red. Además, incluya un nodo de demanda ficticio que reciba (sin costo) la capacidad no utilizada por los proveedores.
C ***c***) Formule y resuelva el modelo en una hoja de cálculo.
C ***d***) Utilice una computadora para resolver el problema sin usar Excel.

D **9.7-1.** Considere el problema del flujo de costo mínimo que se presenta en la siguiente figura, donde los valores de b_i (flujos netos generados) están dados en los nodos, los valores de c_{ij} (costo por unidad de flujo) están dados en los arcos y los valores de u_{ij} (capacidades de los arcos) se encuentran entre los nodos C y D. Realice esta tarea de manera manual.

a) Obtenga una solución BF inicial mediante la solución del árbol de expansión factible con los arcos básicos $A \to B$, $C \to E$, $D \to E$ y $C \to A$ (un arco invertido), en el cual uno de los arcos



no básicos ($C \to B$) también es un arco invertido. Muestre la red resultante (incluya b_i, c_{ij} y u_{ij}) en el mismo formato que el anterior —excepto que debe usar líneas punteadas para dibujar los arcos no básicos—, y agregue los flujos entre paréntesis que están junto a los arcos.

b) Utilice la prueba de optimalidad para verificar que esta solución BF inicial es óptima y que existen múltiples soluciones óptimas. Realice una iteración del método símplex de redes para encontrar la otra solución BF óptima y use estos resultados para identificar las otras soluciones óptimas que no son soluciones BF.
c) Ahora considere la siguiente solución BF.

Arco básico	Flujo	Arco no básico
$A \to D$	20	$A \to B$
$B \to C$	10	$A \to C$
$C \to E$	10	$B \to D$
$D \to E$	20	

A partir de esta solución BF, aplique una iteración del método símplex de redes. Identifique el arco básico entrante, el que sale y la siguiente solución BF, pero no continúe.

9.7-2. Reconsidere el problema del flujo de costo mínimo que se presentó en el problema 9.6-2.

a) Obtenga una solución BF inicial con la solución del árbol de expansión factible con los arcos básicos $A \to B$, $A \to C$, $A \to F$, $B \to D$ y $E \to F$, donde dos de los arcos no básicos ($E \to C$ y $F \to D$) son arcos invertidos.

D, I ***b***) Use el método símplex de redes —puede utilizar la rutina interactiva del IOR Tutorial— para resolver este problema.

9.7-3. Reconsidere el problema de flujo de costo mínimo formulado en el problema 9.6-3

a) Obtenga una solución BF inicial al resolver el árbol de expansión factible sólo mediante las dos vías y la fábrica 1 que remite unidades al almacén 2 a través del centro de distribución.

D, I ***b***) Use el método símplex de redes —puede utilizar la rutina interactiva del IOR Tutorial— para resolver este problema.

D, I **9.7-4.** Reconsidere el problema de flujo de costo mínimo que se formuló en el problema 9.6-4. Comience con la solución BF inicial que corresponde a reemplazar el tractor cada año; utilice el método símplex de redes —puede utilizar la rutina interactiva del IOR Tutorial— para resolver este problema.

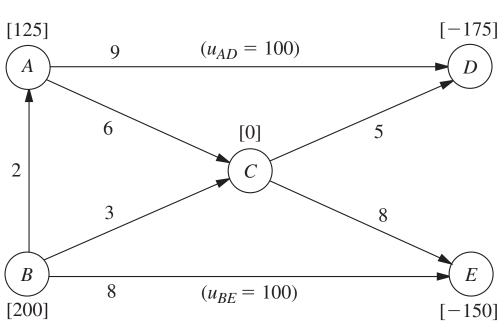
D, I **9.7-5.** En el caso del problema de transporte de la P&T Co., que se presentó en la tabla 8.2, considere su representación de redes como el problema de flujo de costo mínimo de la figura 8.2. Utilice la regla de la esquina noroeste para obtener una solución BF inicial a partir de la tabla 8.2. Use el método símplex de redes —puede utilizar la rutina interactiva en el IOR Tutorial— para resolver este problema (verifique la solución óptima que se proporciona en la sección 8.1).

9.7-6. Considere el problema de transporte del Metro Water District que se presentó en la tabla 8.12.

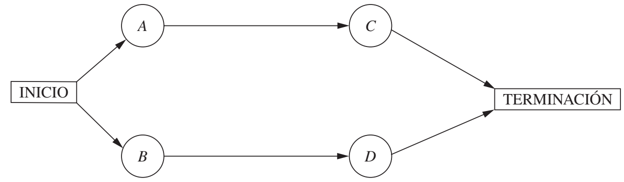
a) Formule la representación de redes para este problema como un problema del flujo de costo mínimo. (*Sugerencia*: Deben eliminarse los arcos para los cuales el flujo está prohibido.)

D, I ***b***) Comience con la solución BF inicial que se dio en la tabla 8.19 y use el método símplex de redes para resolver este problema. Compare la secuencia de soluciones BF que se obtuvieron con la secuencia que resultó de la aplicación del método símplex de transporte en la tabla 8.23.

D, I **9.7-7.** Considere el problema del flujo de costo mínimo que se muestra a continuación, donde los valores de b_i están junto a los nodos, los valores de c_{ij} están junto a los arcos y las u_{ij} finitas están entre paréntesis junto a los arcos. Obtenga una solución BF inicial al resolver el árbol de expansión factible con los arcos básicos $A \to C$, $B \to A$, $C \to D$ y $C \to E$, donde uno de los arcos no básicos ($D \to A$) es un arco *inverso*. Después use el método símplex de redes para resolver este problema (puede utilizar la rutina interactiva del IOR Tutorial).



9.8-1. La compañía Tinker Construction está lista para comenzar un proyecto que debe completarse en 12 meses. Este proyecto tiene cuatro actividades (*A*, *B*, *C*, *D*) en la red de proyecto que se muestra en seguida.



El gerente de proyecto, Sean Murphy, ha concluido que no puede cumplir con la fecha límite si desarrolla todas estas actividades de manera normal. Por tanto, ha decidido utilizar el método CPM de trueques entre tiempo y costo para determinar la manera más económica de acelerar el proyecto para cumplir con la fecha límite. Sean ha recopilado los siguientes datos de las cuatro actividades.

Actividad	Tiempo normal	Tiempo de quiebre	Costo normal	Costo de quiebre
A	8 meses	5 meses	$25 000	$40 000
B	9 meses	7 meses	$20 000	$30 000
C	6 meses	4 meses	$16 000	$24 000
D	7 meses	4 meses	$27 000	$45 000

Use el análisis de costo marginal para resolver este problema.

9.8-2. Reconsidere el problema de la Tinker Construction Co., que se presentó en el problema anterior. Mientras estaba en la universidad, Sean Murphy tomó un curso de IO que dedicó un mes a programación lineal. Por eso ha decidido emplear programación lineal para analizar este problema.

a) Considere la ruta superior a través de la red del proyecto. Formule un modelo de programación lineal de dos variables para resolver el problema de cómo minimizar el costo de realizar esta secuencia de actividades en 12 meses. Utilice el método gráfico para resolver este modelo.

b) Repita el inciso *a*) para la ruta inferior a través de la red del proyecto.

c) Combine los modelos de los incisos *a*) y *b*) en un solo modelo completo de programación lineal para manejar el problema de cómo minimizar el costo de terminar el proyecto en 12 meses. ¿Cuál debería ser una solución óptima de este modelo?

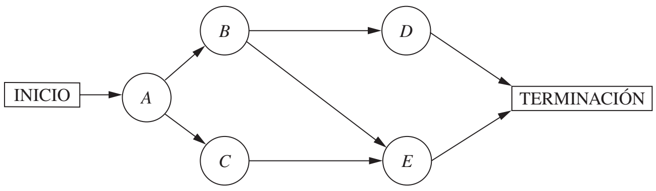
d) Use la formulación de programación lineal del CPM que se presentó en la sección 9.8 para formular un modelo completo de este problema. [Este modelo es un poco más grande que el del inciso *c*) porque el método de formulación también es aplicable a redes de proyecto más complicadas.]

C ***e***) Utilice Excel para resolver este problema.

C ***f***) Utilice otra opción de software para resolver este problema.

C ***g***) Verifique el efecto de cambiar la fecha límite mediante la repetición de los incisos *e*) y *f*) con un límite de 11 meses y después con uno de 13 meses.

9.8-3.* La compañía Good Homes Construction está a punto de comenzar la construcción de una nueva casa muy grande. En la actualidad, el presidente de la compañía, Michael Dean, planea el calendario para llevar a cabo este proyecto. Michael ha identificado las cinco actividades generales (etiquetadas como *A*, *B*, *C*, *D*, *E*) que deberán realizarse de acuerdo con la red de proyecto que se muestra a continuación, seguido por una tabla que proporciona el punto normal y el punto de ruptura para cada una de estas actividades.



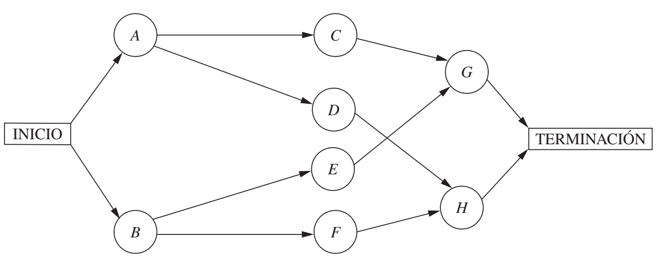
Actividad	Tiempo normal	Tiempo de quiebre	Costo normal	Costo de quiebre
A	3 semanas	2 semanas	$54 000	$60 000
B	4 semanas	3 semanas	$62 000	$65 000
C	5 semanas	2 semanas	$66 000	$70 000
D	3 semanas	1 semana	$40 000	$43 000
E	4 semanas	2 semanas	$75 000	$80 000

Estos costos reflejan los costos directos de la compañía por el material, el equipo y el trabajo directo requeridos para realizar las actividades. Además, la compañía incurre en costos indirectos de proyecto como supervisión y otros indirectos comunes, cargos por intereses sobre el capital, etc. Michael estima que estos costos indirectos serán de 5 000 dólares por semana. Él quiere minimizar el costo global del proyecto. Por lo tanto, para ahorrarse algunos de estos costos indirectos, Michael concluye que debería acortar el proyecto mediante algunas concentraciones hasta el punto en que el costo de aceleración por cada semana adicional eliminada sea menor a 5 000 dólares.

a) Use el análisis de costo marginal para determinar cuáles actividades deberían acelerarse y en cuánto para minimizar el costo global del proyecto. Con este plan, ¿cuál es la duración y el costo de cada actividad? ¿Cuánto dinero ahorra esta aceleración?

C ***b***) Ahora use el enfoque de programación lineal para resolver el inciso *a*), acortando el límite una semana a la vez.

9.8-4. La empresa 21st Century Studios está a punto de iniciar la producción de su película más importante (y más cara) del año. El productor de la película, Dusty Hoffmer, ha decidido utilizar PERT/CPM para planear y controlar este proyecto clave. Dusty ha identificado las ocho actividades más importantes (etiquetadas *A*, *B*, . . . , *H*) que son necesarias para producir la película. Sus relaciones de precedencia se muestran en la red de proyecto de la siguiente figura.



Dusty se ha enterado de que otro estudio también lanzará una película comercial durante la mitad del próximo verano, exactamente cuando la suya iba a ser lanzada, lo cual sería un evento muy desafortunado. Por tanto, él y la alta administración de los 21st Century Studios han concluido que deben acelerar la producción de su película y lanzarla al principio del verano (dentro de 15 semanas) para colocarla como el filme del año. Aunque para alcanzar este objetivo se requerirá de un incremento sustancial en el ya de por sí enorme presupuesto, la administración siente que, si lo logra, obtendrá ganancias en taquilla mucho más grandes a nivel nacional e internacional.

Ahora, Dusty quiere determinar la forma menos costosa de cumplir con la nueva fecha límite, 15 semanas a partir de hoy. Mediante el uso del método CPM de trueque entre tiempo y costo, él ha obtenido los siguientes datos.

Actividad	Tiempo normal	Tiempo de quiebre	Costo normal	Costo de quiebre
A	5 semanas	3 semanas	$24 millones	$36 millones
B	3 semanas	2 semanas	$13 millones	$25 millones
C	4 semanas	2 semanas	$21 millones	$29 millones
D	6 semanas	3 semanas	$30 millones	$50 millones
E	5 semanas	4 semanas	$26 millones	$36 millones
F	7 semanas	4 semanas	$35 millones	$57 millones
G	9 semanas	5 semanas	$30 millones	$53 millones
H	8 semanas	6 semanas	$35 millones	$51 millones

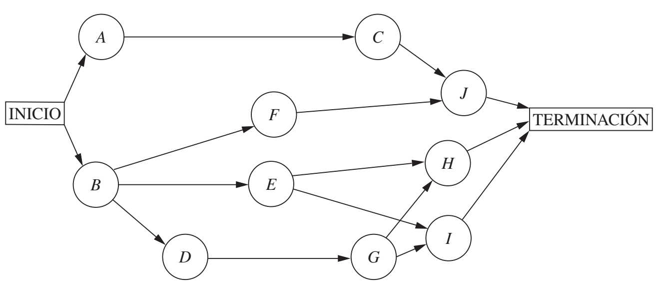
a) Formule un modelo de programación lineal para manejar este problema.

C ***b***) Use Excel para resolver el problema.

C ***c***) Use otra opción de software para resolver el problema.

9.8-5. La compañía Lockhead Aircraft está lista para comenzar un proyecto cuyo objetivo es desarrollar un nuevo avión para la Fuerza Aérea de Estados Unidos. El contrato de la compañía con el Departamento de Defensa impone la conclusión del proyecto en 92 semanas, con penalizaciones impuestas por entregas retrasadas.

El proyecto incluye 10 actividades (etiquetadas *A*, *B*, . . . , *J*), donde sus relaciones de precedencia se muestran en la red de proyecto siguiente



La administración desearía evitar las duras penalizaciones impuestas por no cumplir con la fecha límite establecida en el contrato. Por tanto, ha tomado la decisión de acelerar el proyecto; use el método CPM de trueques entre tiempo y costo para determinar cómo hacerlo en la forma más económica. Los datos que se necesita para aplicar este método se proporcionan a continuación.

Actividad	Tiempo normal	Tiempo de quiebre	Costo normal	Costo de quiebre
A	32 semanas	28 semanas	$160 millones	$180 millones
B	28 semanas	25 semanas	$125 millones	$146 millones
C	36 semanas	31 semanas	$170 millones	$210 millones
D	16 semanas	13 semanas	$ 60 millones	$ 72 millones
E	32 semanas	27 semanas	$135 millones	$160 millones
F	54 semanas	47 semanas	$215 millones	$257 millones
G	17 semanas	15 semanas	$ 90 millones	$ 96 millones
H	20 semanas	17 semanas	$120 millones	$132 millones
I	34 semanas	30 semanas	$190 millones	$226 millones
J	18 semanas	16 semanas	$ 80 millones	$ 84 millones

a) Formule un modelo de programación lineal para manejar este problema.

C *b*) Use Excel para resolverlo.

C *c*) Use otra opción de software para resolver el problema.

9.9-1. De la parte inferior de las referencias proporcionadas al final del capítulo, seleccione una de las aplicaciones de modelos de optimización de redes ganadoras de premios. Lea este artículo y después escriba un resumen de dos páginas acerca de la aplicación y de los beneficios (incluyendo los no financieros) que ésta brinda.

9.9-2. De la parte inferior de las referencias proporcionadas al final del capítulo, seleccione tres de las aplicaciones de modelos de optimización de redes ganadoras de premios. Para cada una de éstas, lea el artículo y después escriba un resumen de una página acerca de la aplicación y de los beneficios (incluyendo los no financieros) que ésta brinda.

CASOS

Caso 9.1 DINERO EN MOVIMIENTO

Jake Nguyen pasa una mano por su cabello antes bien peinado. Se afloja la corbata que tenía un nudo perfecto, y refriega las manos en sus pantalones antes con planchado inmaculado.

Sin duda, hoy no ha sido un buen día.

Durante los últimos meses, Jake ha oído rumores que circulan en Wall Street, provenientes de inversionistas, banqueros y accionistas famosos por su franqueza. Los rumores se refieren a un posible colapso de la economía japonesa; y han corrido sólo como rumores porque creen que si hablan públicamente de sus temores, el colapso puede acelerarse.

Hoy, esos temores se hicieron realidad. Jake y sus colegas se encuentran reunidos alrededor de un pequeño televisor dedicado exclusivamente al canal Bloomberg. Jake mira sin poder creer mientras escucha el horror de lo que ocurre en el mercado japonés, el cual se lleva en su caída al resto de los países de Asia. Se queda helado. Como gerente de inversión extranjera en Asia de Grant Hill Associates, una pequeña empresa de inversión de la costa oeste de Estados Unidos especializada en mercado de dinero, Jake tiene la responsabilidad de enfrentar cualquier efecto negativo del colapso.

Y, por supuesto que Grant Hill experimentará efectos negativos.

Jake no prestó atención a los rumores que advertían sobre un colapso japonés. En su lugar, incrementó de manera considerable la inversión que Grant Hill Associates mantenía en el mercado de Japón. Debido a que ese mercado había tenido mejor desempeño que el esperado durante el año anterior, el mes pasado Jake había aumentado la inversión de 2.5 a 15 millones de dólares en él. En ese momento, el dólar valía 80 yenes.

Ya no. Jake se da cuenta de que la devaluación del yen significa que 1 dólar vale 125 yenes. Podrá liquidar las inversiones que realizó sin perder dinero en yenes, pero la pérdida en dólares al convertir los yenes devaluados será inmensa. Suspira, cierra los ojos y se prepara mentalmente para enfrentar un daño serio.

Su meditación es interrumpida por una voz sonora que lo llama desde la oficina de la esquina. Grant Hill, el presidente de la empresa, le grita: "Nguyen, ¡venga acá de inmediato!"

Jake salta y mira sin entusiasmo hacia la oficina de la esquina en donde está un Grant Hill furioso. Se arregla el pelo, aprieta el nudo de la corbata y camina a la oficina.

Al entrar, Grant Hill lo mira a los ojos y sigue gritando: "¡No quiero que digas una palabra, Nguyen! No hay excusa: ¡sólo arregla el desastre! ¡Saca todo nuestro dinero de Japón! Mi instinto me dice que esto es sólo el principio. Pon el dinero en acciones estadounidenses seguras. ¡Ahora! Y no olvides sacar nuestras posiciones en efectivo de Indonesia y Malasia de una vez."

Jake tiene suficiente sentido común para quedarse callado. Asiente con la cabeza, gira sobre sus talones y casi corre para salir de la oficina.

Una vez a salvo en su escritorio, comienza a formular un plan para mover las inversiones fuera de Japón, Indonesia y Malasia. Su experiencia con la inversión en mercados extranjeros le ha enseñado que cuando se trata de millones de dólares, *cómo* se saca el dinero del mercado extranjero es casi tan importante como *cuándo* se saca. Los banqueros socios de Grant Hill Associates cargan diferentes comisiones por transacciones para convertir el dinero en otra divisa y por enviar cantidades grandes de efectivo alrededor del mundo.

Y ahora, para empeorar las cosas, los gobiernos del oriente asiático han impuesto limitaciones muy estrictas sobre la cantidad de dinero que un individuo o compañía puede cambiar de su moneda a otra moneda extranjera y retirarla del país. El objetivo de esta medida drástica es reducir el flujo de la inversión extranjera saliente para evitar un colapso completo de las economías de la región. Como el dinero en efectivo de Grant Hill Associates llega a 10 500 millones de rupias indonesias y 28 millones de ringgits de Malasia, junto con las inversiones en yenes, no está claro cómo debe convertir este dinero a dólares.

Jake desea encontrar el método más eficiente para convertir estas inversiones a dólares a un mínimo costo. En la página de internet de su compañía siempre puede encontrar las tasas de cambio actualizadas al minuto de la mayoría de las monedas del mundo (tabla 1).

La tabla establece, por ejemplo, que 1 yen japonés es igual a 0.008 dólares. Con unas cuantas llamadas telefónicas obtiene los costos de transacción que debe pagar su compañía por transacciones grandes en estos tiempos críticos (tabla 2).

Jake observa que el cambio de una divisa a otra implica el mismo costo por transacción que la conversión inversa. Por último, Jake encuentra las cantidades máximas de divisas de cada país que su compañía puede convertir en otras divisas en Japón, Indonesia y Malasia (tabla 3).

a) Formule el problema de Jake como un problema de flujo de costo mínimo y trace la red para representarlo gráficamente. Identifique los nodos de recursos y de demanda de la red.

b) ¿Qué transacciones de divisas debe realizar Jake para convertir las inversiones de yenes, rupias y ringgits en dólares estadounidenses para asegurar que Grant Hill Associates tenga la cantidad máxima de dólares después de realizadas las transacciones? ¿Cuánto dinero tiene Jake para invertir en bonos estadounidenses?

c) La Organización Mundial de Comercio prohíbe los límites en las transacciones porque promueven el proteccionismo. Si no existen límites en las transacciones, ¿qué método debe usar Jake para convertir las respectivas divisas de las inversiones asiáticas en dólares?

TABLA 1 Tasas de cambio

De \ A	Yen	Rupia	Ringgit	Dólar estadounidense	Dólar canadiense	Euro	Libra inglesa	Peso
Yen japonés	1	50	0.04	0.008	0.01	0.0064	0.0048	0.0768
Rupia indonesa		1	0.0008	0.00016	0.0002	0.000128	0.000096	0.001536
Ringgit malayo			1	0.2	0.25	0.16	0.12	1.92
Dólar estadounidense				1	1.25	0.8	0.6	9.6
Dólar canadiense					1	0.64	0.48	7.68
Euro						1	0.75	12
Libra inglesa							1	16
Peso mexicano								1

TABLA 2 Costo de transacción, porcentaje

De \ A	Yen	Rupia	Ringgit	Dólar estadounidense	Dólar canadiense	Euro	Libra	Peso
Yen	—	0.5	0.5	0.4	0.4	0.4	0.25	0.5
Rupia		—	0.7	0.5	0.3	0.3	0.75	0.75
Ringgit			—	0.7	0.7	0.4	0.45	0.5
Dólar estadounidense				—	0.05	0.1	0.1	0.1
Dólar canadiense					—	0.2	0.1	0.1
Euro						—	0.05	0.5
Libra							—	0.5
Peso								—

TABLA 3 Límites de transacción en equivalentes a miles de dólares

De \ A	Yen	Rupia	Ringgit	Dólar estadounidense	Dólar canadiense	Euro	Libra inglesa	Peso mexicano
Yen	—	5 000	5 000	2 000	2 000	2 000	2 000	4 000
Rupia	5 000	—	2 000	200	200	1 000	500	200
Ringgit	3 000	4 500	—	1 500	1 500	2 500	1 000	1 000

d) En respuesta al mandato de la Organización Mundial de Comercio que prohíbe imponer límites a las transacciones, el gobierno de Indonesia introduce un nuevo impuesto que provoca un incremento de 500% de los costos de transacciones en el cambio de rupias para proteger su moneda. Dados estos nuevos costos de transacciones, ¿qué operaciones de divisas debe realizar Jake para convertir la inversión en monedas asiáticas en dólares?

e) Jake se da cuenta de que su análisis está incompleto porque no ha incluido todos los aspectos que pueden influir en sus cambios de moneda planeados. Describa otros factores que Jake debe examinar antes de tomar su decisión final.

(*Nota*: Para su conveniencia, en el sitio en internet de este libro se proporciona un archivo de datos para este caso.)

RESUMEN DE LOS CASOS ADICIONALES DE NUESTRO SITIO EN INTERNET (www.mhhe.com/hillier)

Caso 9.2 Ayuda a los aliados

Un ejército rebelde intenta derrocar al gobierno electo de la Federación Rusa. El gobierno de Estados Unidos ha decidido ayudar a su aliado mediante el envío rápido de tropas y suministros. Ahora es necesario desarrollar un plan para embarcar las tropas y los suministros de la manera más eficaz. De acuerdo con la elección de medida global de desempeño, el análisis requiere formular y resolver un problema de la ruta más corta, uno de flujo de costo mínimo o uno de flujo máximo. El análisis subsiguiente requiere la formulación y resolución de un problema de árbol de expansión mínima.

Caso 9.3 Pasos hacia el éxito

La gerencia de una compañía privada ha tomado la decisión de convertirse en pública. En el proceso de realizar la oferta pública inicial de sus acciones, deben completarse muchos pasos relacionados entre sí. La administración desea acelerar este proceso. Por tanto, después de construir una red de proyecto que representa el proceso, aplica el método CPM de trueques entre tiempo y costo.

CAPÍTULO 10

Programación dinámica

La programación dinámica es una técnica matemática útil para la toma de decisiones secuenciales interrelacionadas. Proporciona un procedimiento sistemático para determinar la combinación óptima de decisiones.

En contraste con la programación lineal, no cuenta con una formulación matemática estándar "del" problema de programación dinámica, sino que se trata de un enfoque de tipo general para solucionar problemas; además, las ecuaciones específicas que se usan deben ajustarse a la situación particular. Por tanto, es necesario cierto grado de creatividad y un buen conocimiento de la estructura general de los problemas de programación dinámica para reconocer cuándo y cómo un problema puede ser resuelto por medio de estos procedimientos. Es posible desarrollar mejor estas habilidades mediante la exposición de una gran variedad de aplicaciones de programación dinámica y con el análisis detallado de las características comunes de todas estas situaciones. Con este fin, se presentará un gran número de ejemplos ilustrativos.

10.1 EJEMPLO PROTOTIPO DE PROGRAMACIÓN DINÁMICA

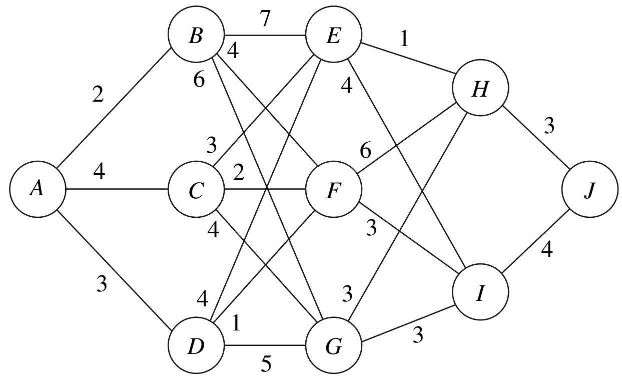
EJEMPLO 1 El problema de la diligencia

El PROBLEMA DE LA DILIGENCIA se construyó[1] especialmente para ilustrar las características e introducir la terminología de la programación dinámica. Este paradigma se refiere a un cazafortunas mítico de Missouri que decide ir al oeste a sumergirse en la fiebre del oro que surgió en California a mediados del siglo XIX. Tiene que hacer el viaje en diligencia a través de territorios sin ley, donde existen serios peligros de ser atacado por merodeadores. A pesar de que su punto de partida y su destino son fijos, tiene muchas opciones en cuanto a qué estados —o territorios— debe elegir como puntos intermedios. En la figura 10.1 se muestran las rutas posibles, en donde cada estado se representa mediante un círculo con una letra; además, en el diagrama, la dirección del viaje es siempre de izquierda a derecha. Como se puede observar, se requieren cuatro etapas —jornadas en diligencia— para viajar desde su punto de partida en el estado A (Missouri) a su destino en el estado J (California).

Este cazafortunas es un hombre prudente preocupado por su seguridad. Después de reflexionar un poco ideó una manera bastante ingeniosa para determinar la ruta más segura. Se ofrecen pólizas de seguros de vida a los pasajeros. Como el costo de la póliza de cualquier jornada en la diligencia está basado en una evaluación cuidadosa de la seguridad del recorrido, la ruta más segura debe ser aquella cuya póliza represente el menor costo total.

El costo de la póliza estándar del viaje en diligencia, del estado i al estado j, que se denota como c_{ij}, es

[1] Este problema fue desarrollado por el profesor Harvey M. Wagner cuando estaba en Stanford University.



■ **FIGURA 10.1**
Sistema de caminos y costos del problema de la diligencia.

	B	*C*	*D*
A	2	4	3

	E	*F*	*G*
B	7	4	6
C	3	2	4
D	4	1	5

	H	*I*
E	1	4
F	6	3
G	3	3

	J
H	3
I	4

Estos costos se muestran en la figura 10.1.

La atención se centrará en la pregunta: ¿Cuál es la ruta que minimiza el costo total de la póliza?

Solución del problema

Observe primero que el procedimiento poco inteligente de elegir la ruta más barata en cada etapa sucesiva no conduce a una decisión óptima global. En caso de adoptar esta estrategia, se obtiene la ruta $A \to B \to F \to I \to J$, con un costo total de 13. Sin embargo, un pequeño sacrificio en una etapa puede permitir mayores ahorros más adelante. Por ejemplo, $A \to D \to F$ es, en total, más barato que $A \to B \to F$.

Un enfoque posible para resolver este problema es el de prueba y error.[2] Sin embargo, el número de rutas posibles es grande (18) y el cálculo del costo total de cada ruta no es una tarea atractiva.

Por fortuna, la programación dinámica proporciona una solución con mucho menos esfuerzo que la enumeración exhaustiva. (El ahorro computacional es enorme cuando se trata de versiones más grandes de este problema.) La programación dinámica comienza con una pequeña porción del problema original y encuentra la solución óptima para este problema pequeño. Después agranda de manera gradual el problema y encuentra la solución óptima actual a partir de la que le precede, hasta resolver el problema original completo.

En el caso del problema de la diligencia, se comienza con el problema sencillo que plantea que el cazafortunas casi ha llegado al final de su viaje y sólo tiene una etapa más —una jornada en la diligencia— por recorrer. La solución óptima obvia de este problema reducido es ir del estado actual —el que sea en el que se encuentre— a su destino final (estado *J*). En cada una de las iteraciones siguientes, el problema aumenta de una en una el número de etapas que le quedan por recorrer para completar el viaje. En cada problema aumentado se puede encontrar la solución óptima del lugar al que debe dirigirse desde cada estado posible, donde se toman en cuenta los resultados obtenidos en la iteración anterior. A continuación se describen los detalles de este procedimiento.

---

[2] Este problema también puede formularse como un *problema de la ruta más corta* (vea la sección 9.3), en el cual los *costos* tienen el papel de las *distancias* en el problema de la ruta más corta. En realidad, el algoritmo de la sección 9.3 utiliza la filosofía de programación dinámica. Sin embargo, debido a que el problema en estudio tiene un número fijo de etapas, el enfoque de programación dinámica que se presentó es el mejor.

Formulación. Sean x_n ($n = 1, 2, 3, 4$) las variables de decisión que representan el destino inmediato de la etapa n (el n-ésimo viaje que se hará en diligencia). En este caso, la ruta seleccionada es $A \to x_1 \to x_2 \to x_3 \to x_4$, donde $x_4 = J$.

Sea $f_n(s, x_n)$ el costo total de la mejor *política* global para enfrentar las etapas *restantes*, mientras el agente de ventas se encuentra en el estado s, listo para iniciar la etapa n y elige x_n como destino inmediato. Dados s y n, sea x_n^* el valor de x_n —no necesariamente único— que minimiza $f_n(s, x_n)$, y sea $f_n^*(s)$ el valor mínimo correspondiente de $f_n(s, x_n)$. Entonces,

$$f_n^*(s) = \min_{x_n} f_n(s, x_n) = f_n(s, x_n^*),$$

donde

$$\begin{aligned} f_n(s, x_n) &= \text{costo inmediato (etapa } n) + \text{costo futuro mínimo (etapas } n + 1 \text{ en adelante)} \\ &= c_{sx_n} + f_{n+1}^*(x_n). \end{aligned}$$

El valor de c_{sx_n} está dado por las tablas anteriores para c_{ij} al establecer $i = s$ (el estado actual) y $j = x_n$ (el destino inmediato). Como el destino final (estado J) se alcanza al terminar la etapa 4, $f_5^*(J) = 0$.

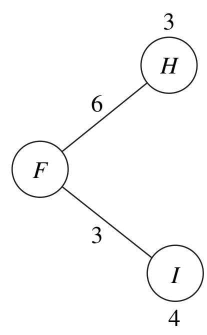
El objetivo es encontrar $f_1^*(A)$ y la ruta correspondiente. La programación dinámica la encuentra al determinar en forma sucesiva $f_4^*(s), f_3^*(s), f_2^*(s)$, para cada uno de los estados posibles s y usar después $f_2^*(s)$ para encontrar $f_1^*(A)$.[3]

Procedimiento de solución. Cuando el cazafortunas tiene sólo una etapa por recorrer ($n = 4$), su ruta de ahí en adelante está perfectamente determinada por su estado actual s (ya sea H o I), así como su destino final, $x_4 = J$, de manera que la ruta de esta última jornada en diligencia es $s \to J$. Por tanto, $f_4^*(s) = f_4(s, J) = c_{s,J}$, la solución inmediata al problema para $n = 4$ es

***n* = 4:**

s	$f_4^*(s)$	x_4^*
H	3	J
I	4	J

Cuando el cazafortunas tiene dos etapas por recorrer ($n = 3$), el procedimiento de solución requiere unos cuantos cálculos. Por ejemplo, suponga que se encuentra en el estado F. Entonces, como se describe en el diagrama, debe ir al estado H o al estado I con unos costos inmediatos respectivos de $c_{F,H} = 6$ o $c_{F,I} = 3$. Si elige el estado H, el costo adicional mínimo al llegar ahí se presenta en la tabla anterior como $f_4^*(H) = 3$, como se muestra sobre el nodo H del diagrama. En consecuencia, el costo total de esta decisión es $6 + 3 = 9$. Si en su lugar elige el estado I, el costo total es $3 + 4 = 7$, que es menor. Por tanto, la opción óptima es esta última, $x_3^* = I$, puesto que proporciona el costo mínimo, $f_3^*(F) = 7$.



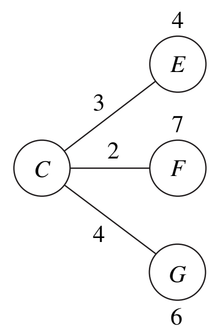
[3] Debido a que el procedimiento implica un recorrido *hacia atrás* etapa por etapa, algunos autores cuentan n también hacia atrás para denotar el número de *etapas que faltan* para llegar al destino. Para simplificar, aquí se usa la forma natural de *contar hacia adelante*.

Son necesarios cálculos similares cuando se parte de los otros dos estados posibles $s = E$ y $s = G$ con dos jornadas por delante. Intente obtener la respuesta con la ayuda tanto de un diagrama (figura 10.1) como del álgebra [combine los valores de c_{ij} y $f_4^*(s)$], para verificar los siguientes resultados del problema con $n = 3$.

$n = 3$:

s \ x_3	$f_3(s, x_3) = c_{sx_3} + f_4^*(x_3)$: H	I	$f_3^*(s)$	x_3^*
E	4	8	4	H
F	9	7	7	I
G	6	7	6	H

La solución de la segunda etapa ($n = 2$), cuando quedan tres jornadas por recorrer, se obtiene en forma parecida. En este caso, $f_2(s, x_2) = c_{sx_2} + f_3^*(x_2)$. Por ejemplo, suponga que el cazafortunas se encuentra en el estado C, como se muestra en el siguiente diagrama.



Ahora deberá ir al estado E, F o G con costos inmediatos respectivos de $c_{C,E} = 3$, $c_{C,F} = 2$ o $c_{C,G} = 4$. Al llegar a este punto, el costo adicional mínimo hasta llegar al destino se presenta en la tabla de $n = 3$ como $f_3^*(E) = 4$, $f_3^*(F) = 7$ o $f_3^*(G) = 6$, respectivamente, como muestra el número colocado arriba de los estados E, F y G en el diagrama anterior. A continuación encontrará un resumen con los cálculos que resultan de las tres alternativas.

$$x_2 = E: \quad f_2(C, E) = c_{C,E} + f_3^*(E) = 3 + 4 = 7.$$
$$x_2 = F: \quad f_2(C, F) = c_{C,F} + f_3^*(F) = 2 + 7 = 9.$$
$$x_2 = G: \quad f_2(C, G) = c_{C,G} + f_3^*(G) = 4 + 6 = 10.$$

El mínimo de estos tres números es 7, por lo que el costo total mínimo desde el estado C al final es $f_2^*(C) = 7$, y el destino inmediato debe ser $x_2^* = E$.

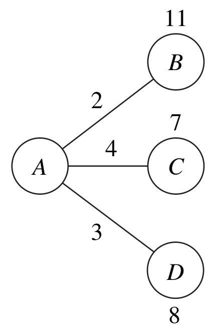
Al hacer cálculos similares cuando se parte del estado B o D (intente esto) se llega a los siguientes resultados para el problema de $n = 2$:

$n = 2$:

s \ x_2	$f_2(s, x_2) = c_{sx_2} + f_3^*(x_2)$: E	F	G	$f_2^*(s)$	x_2^*
B	11	11	12	11	E o F
C	7	9	10	7	E
D	8	8	11	8	E o F

En el primer y tercer renglones de esta tabla observe que E y F empatan como el valor que minimiza x_2, de manera que el destino inmediato desde cualquiera de los estados B o D debe ser $x_2^* = E$ o F.

Si se pasa al problema de la primera etapa ($n = 1$), con las cuatro etapas por recorrer, los cálculos son parecidos a los que se acaban de mostrar para la segunda etapa ($n = 2$), excepto que ahora sólo hay *un* inicio posible, $s = A$, como se muestra en el siguiente diagrama.



A continuación se resumen estos cálculos sobre los tres destinos inmediatos posibles:

$x_1 = B$: $f_1(A, B) = c_{A,B} + f_2^*(B) = 2 + 11 = 13.$
$x_1 = C$: $f_1(A, C) = c_{A,C} + f_2^*(C) = 4 + 7 = 11.$
$x_1 = D$: $f_1(A, D) = c_{A,D} + f_2^*(D) = 3 + 8 = 11.$

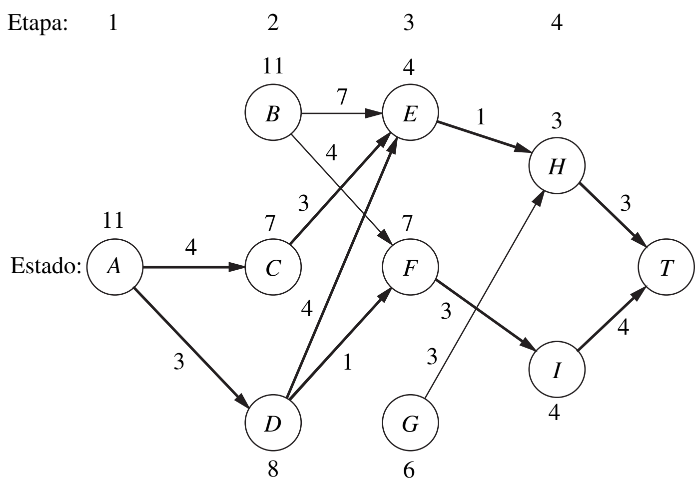
Como el mínimo es 11, $f_1^*(A) = 11$ y $x_1^* = C$ o D, como se muestra en la siguiente tabla:

$n = 1$:

s \ x_1	$f_1(s, x_1) = c_{sx_1} + f_2^*(x_1)$				
	B	**C**	**D**	$f_1^*(s)$	x_1^*
A	13	11	11	11	C o D

Ahora es posible identificar una solución óptima a partir de las cuatro tablas. Los resultados del problema con $n = 1$ indican que el cazafortunas debe elegir como primer destino inmediato el estado C o el estado D. Suponga que elige $x_1^* = C$. Con $n = 2$, el resultado de $s = C$ es $x_2^* = E$. Esto conduce al problema de $n = 3$, que resulta en $x_3^* = H$ con $s = E$, y el problema con $n = 4$ indica que $x_4^* = J$ con $s = H$. Por tanto, una ruta óptima es $A \to C \to E \to H \to J$. Si se elige $x_1^* = D$, se obtienen otras dos rutas óptimas $A \to D \to E \to H \to J$ y $A \to D \to F \to I \to J$. Todas tienen un costo total de $f_1^*(A) = 11$.

En la figura 10.2 se resumen estos resultados del análisis de programación dinámica. Observe que las dos flechas de la etapa 1 se obtienen de la primera y última columnas de la tabla de $n = 1$ y el costo se encuentra en la penúltima columna. Cada una de las otras flechas (y el costo resultante) se lee en un renglón de cada una de las otras tablas, exactamente de la misma manera.



■ **FIGURA 10.2**
Descripción gráfica de la solución de programación dinámica del problema de la diligencia. Cada flecha muestra una política de decisión óptima (el mejor destino inmediato) desde ese estado, donde el número del lado del estado es el costo de ahí hasta el final. Al seguir las flechas oscuras desde *A* hasta *T* se obtienen las tres soluciones óptimas (las rutas que proporcionan el costo mínimo total de 11).

En la siguiente sección se verá que los términos especiales que describen el contexto particular de este problema —*etapa, estado* y *política*— en realidad son parte de la terminología general de programación dinámica con una interpretación análoga en otros contextos.

10.2 CARACTERÍSTICAS DE LOS PROBLEMAS DE PROGRAMACIÓN DINÁMICA

El problema de la diligencia es un prototipo literal de los problemas de programación dinámica. En realidad, el ejemplo se diseñó así, con el propósito de disponer de una interpretación física literal de la estructura abstracta de estos problemas. Por tanto, una manera de reconocer una situación que se puede formular como un problema de programación dinámica es poder identificar una estructura análoga a la del problema de la diligencia.

A continuación se presentan y estudian estas características básicas que distinguen a los problemas de programación dinámica.

1. El problema se puede dividir en **etapas**, cada una de las cuales requiere de una **política de decisión**.

   En el problema de la diligencia se hizo una división literal en cuatro etapas (viajes) que corresponden a las cuatro jornadas en diligencia. La política de decisión en cada etapa fue qué póliza de seguro elegir, esto es, qué destino elegir para la siguiente jornada en diligencia. De manera parecida, otros problemas de programación dinámica requieren tomar una *serie de decisiones interrelacionadas*, cada una de las cuales corresponde a una etapa del problema.
2. Cada etapa tiene cierto número de **estados** asociados con su inicio.

   Los estados asociados con cada etapa del problema de la diligencia son los estados (o territorios) en los que el cazafortunas puede encontrarse al iniciar esa jornada específica del viaje. En general, los estados son las distintas *condiciones posibles* en las que se puede encontrar el sistema en cada etapa del problema. El número de estados puede ser finito —como en el problema de la diligencia— o infinito, como en otros ejemplos subsecuentes.
3. El efecto de la política de decisión en cada etapa es *transformar el estado actual en un estado asociado con el inicio de la siguiente etapa*, quizá según una distribución de probabilidad.

   La decisión del cazafortunas sobre su siguiente destino lo conduce de su estado actual al siguiente estado en su viaje. Este procedimiento sugiere que los problemas de programación dinámica se pueden interpretar en términos de las *redes* descritas en el capítulo 9. Cada *nodo* corresponde a un *estado*. La red consistiría en columnas de nodos, donde cada *columna* corresponde a una *etapa*, en forma tal que el flujo que sale de un nodo sólo puede ir a un nodo de la siguiente columna a la derecha. El valor asignado a cada rama que conecta dos nodos puede interpretarse algunas veces como la *contribución inmediata* a la función objetivo que se obtiene al tomar esa política de decisión. En la mayor parte de los casos, el objetivo corresponde a encontrar la trayectoria *más corta* o bien la *más larga* a través de la red.
4. El procedimiento de solución está diseñado para encontrar una **política óptima** para manejar el problema completo, es decir, una receta para elaborar la política de decisión óptima para cada etapa en *cada* uno de los estados posibles.

   En el problema de la diligencia, el procedimiento de solución se basa en construir una tabla de cada etapa (n) que prescribe la decisión óptima (x_n^*) para cada estado posible (s). Así, además de identificar las tres *soluciones óptimas* (rutas óptimas) del problema completo, los resultados muestran también cómo debe proceder el cazafortunas en caso de que sea desviado a un estado que no se encuentra en la ruta óptima. En cualquier problema, la programación dinámica proporciona este tipo de receta *política* sobre qué hacer en todas las circunstancias posibles (a esto se debe que la decisión real que se toma al llegar a un estado en particular se llama *política* de decisión). Proporcionar esta información adicional, en vez de sólo especificar una solución óptima —secuencia óptima de decisiones—, puede ser muy valioso en muchas situaciones que incluyen el análisis de sensibilidad.
5. Dado el estado actual, una *política óptima para las etapas restantes* es *independiente* de la política adoptada en *etapas anteriores*. Por tanto, la decisión inmediata óptima depende sólo del estado actual y no de cómo se llegó ahí. Éste es el **principio de optimalidad** de la programación dinámica.

Dado el estado en el que se localiza el cazafortunas, la póliza de seguro de vida óptima (y su ruta asociada) desde este lugar en adelante es independiente de cómo llegó ahí. En general, en los problemas de programación dinámica, el conocimiento del estado actual del sistema expresa toda la información sobre su comportamiento anterior, información que es necesaria para determinar la política óptima de ahí en adelante. (Esta propiedad es la *propiedad markoviana* que se presentará en la sección 16.2.) Un problema que carezca de esta propiedad no se puede formular como un problema de programación dinámica.

6. El procedimiento de solución comienza cuando se determina la *política óptima para la última etapa*.

La política óptima para la última etapa prescribe la política óptima de decisión para *cada* estado posible en esa etapa. Es común que la decisión de este problema de una etapa sea trivial, como lo fue en el problema de la diligencia.

7. Se dispone de una **relación recursiva** que identifica la política óptima para la etapa n, dada la política óptima para la etapa $n + 1$.

En el problema de la diligencia, la relación recursiva que se obtuvo es

$$f_n^*(s) = \min_{x_n} \{c_{sx_n} + f_{n+1}^*(x_n)\}.$$

Entonces, para encontrar la *política óptima de decisión* cuando se comienza en el estado s de la etapa n se necesita encontrar el valor que minimice x_n. El costo mínimo correspondiente se obtiene al usar este valor de x_n para después seguir la política óptima cuando el proceso se encuentra en el estado x_n en la etapa $n + 1$.

La forma precisa de la relación recursiva difiere de un problema a otro de programación dinámica, pero se usará una notación análoga a la que se introdujo en la sección anterior, como se resume a continuación:

N = número de etapas.
n = etiqueta de la etapa actual ($n = 1, 2, \ldots, N$).
s_n = *estado* actual de la etapa n.
x_n = variable de decisión de la etapa n.
x_n^* = valor óptimo de x_n (dado s_n).
$f_n(s_n, x_n)$ = contribución de los estados $n, n + 1, \ldots, N$ a la función objetivo si el sistema se encuentra en el estado s_n en la etapa n, la decisión inmediata es x_n, y en adelante se toman decisiones óptimas.

$f_n^*(s_n) = f_n(s_n, x_n^*)$.

La relación recursiva siempre tendrá la forma

$$f_n^*(s_n) = \max_{x_n} \{f_n(s_n, x_n)\} \qquad \text{o} \qquad f_n^*(s_n) = \min_{x_n} \{f_n(s_n, x_n)\},$$

donde $f_n(s_n, x_n)$ se escribe en términos de s_n, x_n, $f_{n+1}^*(s_{n+1})$ y tal vez alguna medida de la contribución inmediata de x_n a la función objetivo. Lo que hace que la expresión de $f_n^*(s_n)$ sea una relación recursiva es la inclusión de $f_{n+1}^*(s_{n+1})$, en el lado derecho, de manera que $f_n^*(s_n)$ está definida en términos de $f_{n+1}^*(s_{n+1})$.

La relación recursiva recurre constantemente a las etapas posteriores a medida que se trabaja hacia atrás una etapa a la vez. Cuando el número de la etapa actual n disminuye su valor en 1, la nueva función $f_n^*(s_n)$ se obtiene mediante el uso de la función $f_{n+1}^*(s_{n+1})$ que se obtuvo en la iteración anterior; después, el proceso se repite cada nueva iteración. Esta propiedad se refuerza en la siguiente (y última) característica de programación dinámica.

8. Cuando se usa esta relación recursiva, el procedimiento de solución comienza al final y se mueve *hacia atrás* etapa por etapa para encontrar cada vez la política óptima para esa etapa hasta que encuentra la política óptima desde la etapa *inicial*. Esta política óptima lleva de inmediato a una solución óptima para el problema completo, es decir, x_1^* para el estado inicial s_1, después x_2^* para el estado s_2 que resulta, luego x_3^* para el estado s_3 que se obtiene, y así sucesivamente hasta x_N^* para el estado s_N resultante.

Este movimiento hacia atrás se mostró en el problema de la diligencia, en el que se encontró la política óptima, en forma sucesiva, donde cada estado se iniciaba de las etapas respec-

tivas 4, 3, 2 y 1.[4] Para todos los problemas de programación dinámica, se obtiene una tabla como la siguiente para cada etapa ($n = N, N - 1, \dots, 1$).

s_n \ x_n	$f_n(s_n, x_n)$	$f_n^*(s_n)$	x_n^*

Cuando se obtiene esta tabla de la etapa inicial ($n = 1$), el problema queda resuelto. Como se conoce el estado de la etapa inicial, la primera decisión está especificada por x_1^* en esta tabla. El valor óptimo de las otras variables de decisión queda, a su vez, especificado por las otras tablas según el estado del sistema que se obtiene al tomar la decisión anterior.

10.3 PROGRAMACIÓN DINÁMICA DETERMINÍSTICA

Esta sección profundiza en el enfoque de programación dinámica en los problemas *determinísticos*, en los cuales el *estado* de la siguiente *etapa* está *determinado por completo* por el *estado* y la *política de decisión* de la *etapa actual.* El caso *probabilístico* en el que existe una distribución de probabilidad del valor posible del siguiente estado se analizará en la sección siguiente.

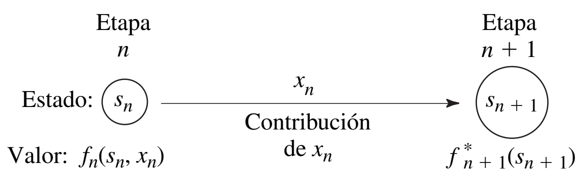
La programación dinámica determinística se puede describir en un diagrama como el de la figura 10.3. En la etapa n el proceso está en algún estado s_n. Al tomar la decisión x_n se mueve a algún estado s_{n+1} en la etapa $n + 1$. La contribución a la función objetivo de ese punto *en adelante* se calculó como $f_{n+1}^*(s_{n+1})$. La política de decisión x_n también contribuye a la función objetivo. Al combinar estas dos cantidades en la forma apropiada se obtiene $f_n(s_n, x_n)$, la contribución de la etapa n en adelante. De la optimización respecto de x_n se obtiene entonces $f_n^*(s_n) = f_n(s_n, x_n^*)$. Una vez determinados x_n^* y $f_n^*(s_n)$ para cada valor posible de s_n, el procedimiento de solución se mueve hacia atrás una etapa.

Una manera de clasificar los problemas de programación dinámica determinística es con base en la *forma de la función objetivo.* Por ejemplo, el objetivo puede ser minimizar la suma de las contribuciones en cada etapa individual —como en el problema de la diligencia—, o maximizar esa suma, o bien minimizar el producto de los términos, etc. Otra clasificación se puede hacer en términos de la naturaleza del *conjunto de estados* en las respectivas etapas. En particular, los estados s_n pueden estar representados por una variable de estado *discreta* —como en el problema de la diligencia—, o por una variable de estado *continua*, o tal vez se requiera un *vector* de estado (más de una variable). De manera similar, las variables de decisión ($x_1, x_2, \dots, x_N$) también pueden ser discretas o continuas.

Se presentarán varios ejemplos para ilustrar estas posibilidades, pero es más importante el hecho de que dichos ejemplos ponen de manifiesto que estas diferencias, en apariencia grandes, en realidad son intrascendentes —excepto en términos de la dificultad de los cálculos— pues la estructura básica de la figura 10.3 permanece igual.

El primer nuevo ejemplo surge en un contexto muy distinto al del problema de la diligencia, pero tiene la misma *formulación matemática*, aunque esta vez se trata de *maximizar* en lugar de minimizar una suma.

FIGURA 10.3
Estructura básica de programación dinámica determinística.



[4] En realidad, el procedimiento de solución de este problema se puede mover hacia atrás o hacia adelante. Sin embargo, en muchos problemas —en particular cuando las etapas corresponden al *tiempo*—, el procedimiento debe moverse hacia atrás.

Recuadro de aplicación

Seis días después que Saddam Hussein ordenó a sus fuerzas militares iraquíes la invasión de Kuwait el 2 de agosto de 1990, Estados Unidos comenzó un largo proceso consistente en enviar un gran número de unidades militares y cargamento a la región. Después de desarrollar una fuerza de coalición conformada por 35 naciones, se inició la operación militar llamada **Tormenta del Desierto** el 17 de enero de 1991, con el fin de expulsar a las tropas iraquíes de Kuwait. Esta decisión condujo a una rápida victoria por parte de las fuerzas de coalición, que abandonaron Kuwait e invadieron Iraq.

El reto desde el punto de vista logístico involucrado en el transporte expedito de las tropas y cargamento necesarios a la zona de conflicto era enorme. Una misión típica de envío de tropas y carga desde Estados Unidos hasta el golfo Pérsico requería de tres días viaje redondo, tocaba siete o más campos de aviación, consumía alrededor de un millón de libras de combustible y costaba 280 000 dólares. Durante dicha operación, el Comando Militar de Transporte realizó más de 100 de estas misiones diariamente, ya que estaba a cargo de la misión más grande de transporte aéreo de la historia.

Para cumplir con este reto, se aplicó la investigación de operaciones con el fin de desarrollar los sistemas de soporte a las decisiones necesarios para programar y enrutar las misiones de transporte aéreo. La técnica de Investigación de Operaciones que se utilizó para manejar este proceso fue la *programación dinámica*. Las etapas en la formulación de la programación dinámica corresponden a los campos aéreos en la red de vuelos relevantes para la misión. En un determinado campo aéreo, los estados están caracterizados por el tiempo de salida del y el trabajo disponible que queda de la tripulación actual. La función objetivo que se debía minimizar era la suma ponderada de varias medidas de desempeño: la tardanza en la entrega, el tiempo de vuelo de la misión, el tiempo que se permanece en tierra y el número de cambios de tripulación. Las restricciones incluían una cota inferior en la carga que transportaría la misión y cotas superiores en la disponibilidad de la tripulación y los recursos de soporte en tierra de los campos aéreos.

Esta aplicación de la programación lineal tuvo un efecto dramático respecto a su habilidad para entregar, de manera rápida, la carga y personal necesarios en el golfo Pérsico para apoyar la Operación Tormenta del Desierto. Por ejemplo, cuando se platicó con los desarrolladores de este método, el comandante en jefe de operaciones y transporte de MAC comentó: "Les garantizo que no lo hubiéramos podido hacer (el despliegue de tropas en el golfo Pérsico) sin su ayuda y las contribuciones que ustedes hicieron (los sistemas de soporte a las decisiones); no lo hubiéramos podido hacer en lo absoluto."

Fuente: M. C. Hilliard, R. S. Solanki, C. Liu, I. K. Busch, G. Harrison y R. D. Kraemer: "Scheduling the Operation Desert Storm Airlift: An advanced Automated Scheduling Support System", en *Interfaces*, **22**(1): 131-146, enero-febrero, 1992.

EJEMPLO 2 Distribución de brigadas médicas entre países

El CONSEJO MUNDIAL DE LA SALUD se dedica a mejorar la atención médica en los países en desarrollo. En la actualidad dispone de cinco brigadas médicas para asignarlas a tres de ellos con el fin de mejorar el cuidado de la salud, la educación para la salud y los programas de capacitación. El consejo debe determinar cuántas brigadas asignar —si lo hace— a cada uno de estos países para maximizar la medida de eficiencia de las cinco brigadas. Éstas deben mantenerse como están constituidas, es decir, el número asignado a cada país debe ser un entero.

La medida de desempeño se evalúa en términos de los *años de vida adicionales por persona.* (En el caso de un país específico, esta medida es igual al *incremento del promedio de vida esperado* en años, multiplicado por su población.) En la tabla 10.1 se proporcionan las estimaciones de estos años de vida adicionales por persona (en múltiplos de mil) para cada país y para cada asignación posible de brigadas médicas.

¿Cuál es la asignación que maximiza la medida de desempeño?

TABLA 10.1 Datos del problema del Consejo Mundial de la Salud

	Miles de años-persona de vida adicionales		
	País		
Brigadas médicas	1	2	3
0	0	0	0
1	45	20	50
2	70	45	70
3	90	75	80
4	105	110	100
5	120	150	130

Formulación. Este problema requiere tomar tres *decisiones interrelacionadas*: cuántas brigadas conviene asignar a cada uno de los tres países. A pesar de que no existe una secuencia fija, estos tres países se pueden considerar como las tres etapas en la formulación de programación dinámica. Las variables de decisión x_n ($n = 1, 2, 3$) son el número de brigadas que se asignan a la etapa (país) n.

La identificación de los estados puede no ser tan evidente. Para determinarlos se hacen preguntas como las siguientes. ¿Qué es lo que cambia de una etapa a la otra? Dado que se han tomado las decisiones en las etapas anteriores, ¿cómo se puede describir el estado de la situación? ¿Qué información sobre el estado actual de las cosas se necesita para determinar la política óptima de aquí en adelante? Con esta base, una opción apropiada para "el estado del sistema" es

s_n = número de brigadas médicas todavía disponibles para ser asignadas a los países restantes ($n, \ldots, 3$).

Así, en la etapa 1 (país 1), cuando todavía quedan por asignar brigadas a los tres países, $s_1 = 5$. Sin embargo, en las etapas 2 o 3 (países 2 o 3), s_n es exactamente 5 menos el número de brigadas asignadas en etapas anteriores, de manera que la secuencia de estados es

$$s_1 = 5, \qquad s_2 = 5 - x_1, \qquad s_3 = s_2 - x_2.$$

Debido al procedimiento de programación dinámica que resuelve hacia atrás etapa por etapa, cuando se trabaja en la etapa 2 o 3 todavía no se han obtenido las asignaciones de las etapas anteriores. Por tanto, se deben considerar todos los estados posibles al iniciar la etapa 2 o 3, es decir, $s_n = 0$, 1, 2, 3, 4 o 5.

En la figura 10.4 se muestran los estados que deben considerarse en cada etapa. Las ligaduras (segmentos de recta) indican las transiciones posibles de estados de una etapa a la siguiente después de hacer una asignación factible de brigadas médicas al país en cuestión. Los números al lado de las ligaduras son las contribuciones correspondientes a la medida de desempeño, los cuales se tomaron de la tabla 10.1. Desde la perspectiva de esta figura, el problema global es encontrar la trayectoria del estado inicial 5 (inicio de la etapa 1) al estado final 0 (después de la etapa 3) que maximice la suma de los números a lo largo de la ruta.

Para establecer el problema completo en forma matemática, sea $p_i(x_i)$ la medida de desempeño que se obtiene si se asignan x_i brigadas médicas al país i, según los datos de la tabla 10.1. Entonces, el objetivo es elegir x_1, x_2 y x_3 para

$$\text{Maximizar} \quad \sum_{i=1}^{3} p_i(x_i),$$

sujeta a

$$\sum_{i=1}^{3} x_i = 5,$$

y

las x_i son enteros no negativos.

Si se usa la notación que se presentó en la sección 10.2, se observa que $f_n(s_n, x_n)$ es

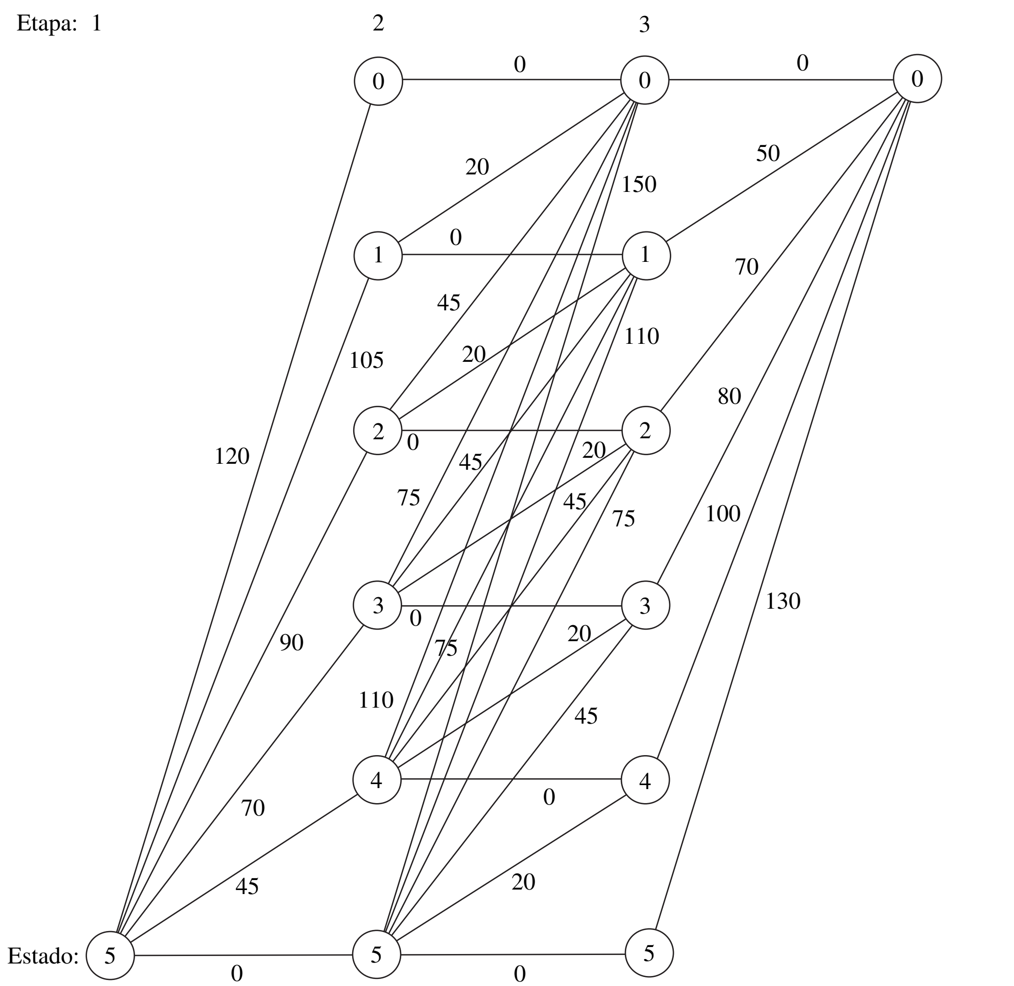
$$f_n(s_n, x_n) = p_n(x_n) + \text{máx} \sum_{i=n+1}^{3} p_i(x_i),$$

donde el máximo se toma sobre las $x_{n+1}, \ldots, x_3$ tales que

$$\sum_{i=n}^{3} x_i = s_n$$

y las x_i son enteros no negativos, para $n = 1, 2, 3$. Además,

$$f_n^*(s_n) = \max_{x_n = 0, 1, \ldots, s_n} f_n(s_n, x_n)$$



■ **FIGURA 10.4**
Descripción gráfica del problema del Consejo Mundial de la Salud que muestra los estados posibles en cada etapa, las transiciones posibles de los estados y las contribuciones correspondientes a la medida de desempeño.

Por lo tanto,

$$f_n(s_n, x_n) = p_n(x_n) + f^*_{n+1}(s_n - x_n)$$

(con f^*_4 definido como cero). En la figura 10.5 se resumen estas relaciones básicas.

En consecuencia, la *relación recursiva* que enlaza las funciones f^*_1, f^*_2 y f^*_3 en este problema es

$$f^*_n(s_n) = \max_{x_n=0,1,\ldots,s_n} \{p_n(x_n) + f^*_{n+1}(s_n - x_n)\}, \qquad \text{para } n = 1, 2.$$

En el caso de la última etapa ($n = 3$),

$$f^*_3(s_3) = \max_{x_3=0,1,\ldots,s_3} p_3(x_3).$$

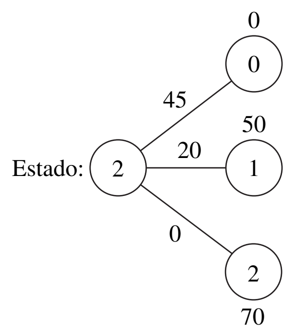
A continuación se presentan los cálculos que resultan de la programación dinámica.

Procedimiento de solución. A partir de la última etapa ($n = 3$), se observa que los valores de $p_3(x_3)$, incluidos en la última columna de la tabla 10.1, aumentan hacia abajo de la columna. Entonces, si se dispone de s_3 brigadas médicas para asignar al país 3, el máximo de $p_3(x_3)$ se logra de manera automática cuando se asignan todas las s_3 brigadas; así, $x^*_3 = s_3$ y $f^*_3(s_3) = p_3(s_3)$, como se puede ver en la siguiente tabla.

***n* = 3:**

s_3	$f_3^*(s_3)$	x_3^*
0	0	0
1	50	1
2	70	2
3	80	3
4	100	4
5	130	5

Ahora el proceso se mueve hacia atrás para comenzar con la penúltima etapa ($n = 2$). Con objeto de determinar x_2^* se necesita calcular y comparar $f_2(s_2, x_2)$ para los distintos valores posibles de x_2, esto es, $x_2 = 0, 1, \ldots, s_2$. La siguiente gráfica ilustra esta situación cuando $s_2 = 2$.



Este diagrama corresponde a la figura 10.5, salvo que ahora se muestran los tres estados posibles en la etapa 3. Así, si $x_2 = 0$, el estado que resulta de la etapa 3 será $s_2 - x_2 = 2 - 0 = 2$, mientras que si $x_2 = 1$ se llega al estado 1 y $x_2 = 2$ conduce al estado 0. Los valores correspondientes de $p_2(x_2)$ en la columna del país 2 de la tabla 10.1 se muestran junto a las ligas y los valores de $f_3^*(s_2 - x_2)$ que provienen de la tabla para $n = 3$ y que se dan junto a los nodos de la etapa 3. Los cálculos requeridos para el caso de $s_2 = 2$ se resumen abajo:

Fórmula: $f_2(2, x_2) = p_2(x_2) + f_3^*(2 - x_2)$.
$p_2(x_2)$ se localiza en la columna del país 2 de la tabla 10.1
$f_3^*(2 - x_2)$ es tomado de $n = 3$ de la tabla anterior.

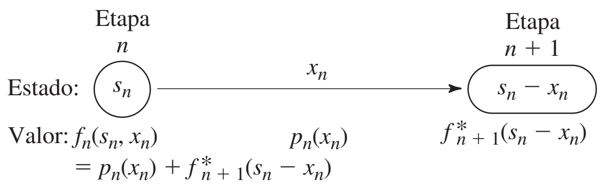
$x_2 = 0$: $f_2(2, 0) = p_2(0) + f_3^*(2) = 0 + 70 = 70$.
$x_2 = 1$: $f_2(2, 1) = p_2(1) + f_3^*(1) = 20 + 50 = 70$.
$x_2 = 2$: $f_2(2, 2) = p_2(2) + f_3^*(0) = 45 + 0 = 45$.

Dado que el objetivo es la *maximización*, $x_2^* = 0$ o 1 con $f_2^*(2) = 70$.

Si se continúa de una manera similar con los otros valores posibles de s_2 (inténtelo) se llega a la siguiente tabla.

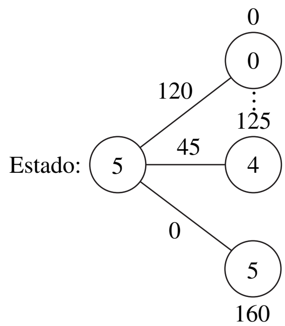
***n* = 2:**

s_2 \ x_2	$f_2(s_2, x_2) = p_2(x_2) + f_3^*(s_2 - x_2)$						$f_2^*(s_2)$	x_2^*
	0	**1**	**2**	**3**	**4**	**5**		
0	0						0	0
1	50	20					50	0
2	70	70	45				70	0 o 1
3	80	90	95	75			95	2
4	100	100	115	125	110		125	3
5	130	120	125	145	160	150	160	4



■ **FIGURA 10.5**
Estructura básica del problema del Consejo Mundial de la Salud.

En este punto, el procedimiento se mueve hacia atrás para resolver el problema original desde el principio, la etapa 1 ($n = 1$). En este caso, el único estado que debe considerarse es el inicial con $s_1 = 5$, como se describe en seguida.



Si se asignan x_1 brigadas médicas al país 1 se llega al estado $5 - x_1$ brigadas disponibles en la etapa 2, con la elección $x_1 = 0$ se llega al nodo inferior de la derecha, $x_1 = 1$ conduce al siguiente nodo hacia arriba, etc., hasta el nodo superior con $x_1 = 5$. Junto a las ligaduras se muestran los valores correspondientes a $p_1(x_1)$ de la tabla 10.1. Los números junto a los nodos se obtienen de la columna de $f_2^*(s_2)$ de la tabla con $n = 2$. Como antes, se resumen los cálculos de cada valor posible de la variable de decisión que implican sumar los valores de la ligadura correspondiente y el valor del nodo:

Fórmula: $f_1(5, x_1) = p_1(x_1) + f_2^*(5 - x_1)$.
$p_1(x_1)$ se proporciona en la columna del país 1 de la tabla 10.1.
$f_2^*(5 - x_1)$ se presenta en la tabla de $n = 2$.

$x_1 = 0$: $f_1(5, 0) = p_1(0) + f_2^*(5) = 0 + 160 = 160$.
$x_1 = 1$: $f_1(5, 1) = p_1(1) + f_2^*(4) = 45 + 125 = 170$.
⋮
$x_1 = 5$: $f_1(5, 5) = p_1(5) + f_2^*(0) = 120 + 0 = 120$.

En la siguiente tabla se muestran los cálculos similares para $x_1 = 2, 3, 4$ (inténtelos) que verifican que $x_1^* = 1$ con $f_1^*(5) = 170$.

$n = 1$:

s_1 \ x_2	$f_1(s_1, x_1) = p_1(x_1) + f_2^*(s_1 - x_1)$							
	0	**1**	**2**	**3**	**4**	**5**	$f_1^*(s_1)$	x_1^*
5	160	170	165	160	155	120	170	1

Así, la solución óptima se obtiene con $x_1^* = 1$, lo que hace que $s_2 = 5 - 1 = 4$, entonces $x_2^* = 3$, lo que hace $s_3 = 4 - 3 = 1$, por tanto $x_3^* = 1$. Como $f_1^*(5) = 170$, esta asignación de (1, 3, 1) brigadas médicas a los tres países conducirá a un total estimado de 170 000 años de vida adicionales, lo que significa por lo menos 5 000 más que para cualquier otra asignación.

Estos resultados del análisis de programación dinámica también se resumen en la figura 10.6.

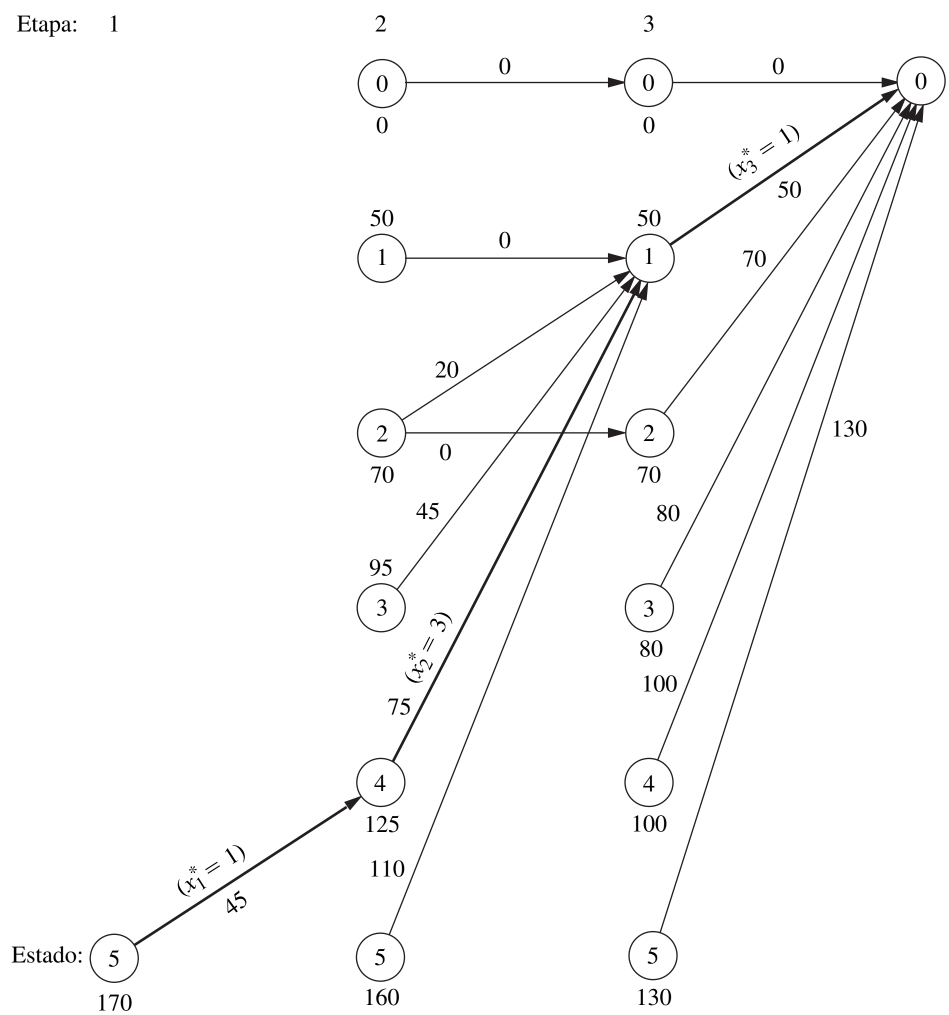


FIGURA 10.6
Descripción gráfica de la solución de programación dinámica del problema del Consejo Mundial de la Salud. Una flecha del estado s_n al estado s_{n+1} indica que una política de decisión óptima desde el estado s_n es asignar $(s_n - s_{n+1})$ brigadas médicas al país n. Al asignar las brigadas de esta manera, si se siguen las flechas oscuras desde el estado inicial, se obtiene la solución óptima.

Un problema de tipo prevalente: problema de distribución de esfuerzo

En el ejemplo anterior se ilustra un tipo bastante común de problemas de programación dinámica llamado *problema de distribución de esfuerzo*. En este grupo de problemas existe sólo una clase de *recurso* que debe asignarse a cierto número de *actividades*. El objetivo es determinar cómo distribuir el esfuerzo (el recurso) entre las actividades de la manera más eficaz. En el caso del Consejo Mundial de la Salud, el recurso de que se trata es el conjunto de brigadas médicas, y las tres actividades son los trabajos sobre el cuidado de la salud en los tres países.

Supuestos. Esta interpretación de distribución de recursos entre actividades debe ser familiar al lector ya que es la más común entre los problemas de programación lineal que se presentaron al principio del capítulo 3. Sin embargo, existen algunas diferencias importantes entre el problema de distribución del esfuerzo y la programación lineal que deben ayudar a aclarar en qué difiere la programación dinámica de otras áreas de programación matemática.

Una diferencia clave es que el problema de distribución del esfuerzo incluye sólo *un recurso* (una restricción funcional), mientras que la programación lineal puede manejar cientos o miles de recursos. (En principio, la programación dinámica puede manejar algo más que un recurso, como se verá en el ejemplo 5, en el que se resuelve el problema de tres recursos de la Wyndor Glass Co., pero pronto comienza a perder eficiencia a medida que aumenta el número de recursos.)

Por otro lado, el problema de distribución del esfuerzo es mucho más general que la programación lineal. Considere los cuatro supuestos de programación lineal que se presentaron en la sec-

ción 3.3: proporcionalidad, aditividad, divisibilidad y certidumbre. La *proporcionalidad* se viola por rutina en casi todos los problemas de programación dinámica, incluso los de distribución de esfuerzo (por ejemplo, la tabla 10.1 viola la proporcionalidad). La *divisibilidad* también se viola con frecuencia, como en el ejemplo 2, en donde las variables de decisión deben ser enteras. En realidad, los cálculos de programación dinámica incrementan su grado de complejidad cuando se cumple la divisibilidad (como en los ejemplos 4 y 5). Aunque en este libro se considerará el problema de distribución de esfuerzo sólo bajo condiciones de *certidumbre*, ésta no es necesaria y muchos otros problemas de programación dinámica tampoco respetan este supuesto (como se describe en la sección 10.4).

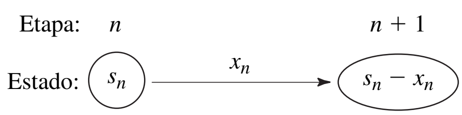
De los cuatro supuestos de programación lineal, el *único* necesario en el problema de distribución de esfuerzo —o en otros problemas de programación dinámica— es el de *aditividad* (o su análogo para funciones que implican el *producto* de términos). Este supuesto se requiere para satisfacer el *principio de optimalidad* de la programación dinámica (característica 5 de la sección 10.2).

Formulación. Debido a que los problemas de distribución de esfuerzo siempre incluyen la asignación de un tipo de recurso a cierto número de actividades, siempre tienen la siguiente formulación de programación dinámica, donde el orden de las actividades es arbitrario:

Etapa n = actividad n ($n = 1, 2, \ldots, N$).
x_n = cantidad de recursos asignados a la actividad n.
Estado s_n = cantidad de recursos todavía disponibles para asignarse a las actividades restantes ($n, \ldots, N$).

La razón para definir el estado s_n de esta manera es que la cantidad de recursos todavía disponibles para asignar es precisamente la información necesaria sobre el estado actual de la situación —al comenzar la etapa n— para tomar las decisiones de asignación para el resto de las actividades.

Cuando el sistema inicia la etapa n en el estado s_n, la elección de x_n siempre da como resultado que el siguiente estado de la etapa $n + 1$ sea $s_{n+1} = s_n - x_n$, como lo muestra el siguiente diagrama:[5]



Observe que la estructura de este diagrama corresponde a la que se muestra en la figura 10.5 en el caso del ejemplo del Consejo Mundial de la Salud de distribución de esfuerzo. Lo que puede diferir de un ejemplo a otro es el *resto* de lo que se muestra en la figura 10.5, es decir, la relación entre $f_n(s_n, x_n)$ y $f^*_{n+1}(s_n - x_n)$, y después la *relación recursiva* que se establece entre las funciones f^*_n y f^*_{n+1}. Estas relaciones dependen de la función objetivo específica del problema.

La estructura del siguiente ejemplo es similar a la del problema del Consejo Mundial de la Salud porque también se trata de un problema de distribución de esfuerzo. Sin embargo, su relación recursiva difiere en que su objetivo es minimizar un producto de términos de las etapas respectivas.

A primera vista, puede parecer que este ejemplo *no* es un problema de programación dinámica determinística pues trata con probabilidades. No obstante, se ajusta a la definición porque en la siguiente etapa el estado queda completamente determinado por el estado y la política de decisión de la etapa actual.

EJEMPLO 3 Distribución de científicos entre grupos de investigación

El grupo encargado de un proyecto espacial del gobierno lleva a cabo una investigación sobre cierto problema de ingeniería que debe resolverse antes de que los seres humanos puedan viajar seguros a Marte. Por el momento, tres equipos de investigación tratan de resolver el problema desde tres puntos de vista diferentes. Se estima que en las circunstancias actuales, la probabilidad de que los

[5] Esta afirmación supone que x_n y s_n están expresados en las mismas unidades. Si es más conveniente definir x_n como si la cantidad del recurso asignado a la actividad n fuese $a_n x_n$, entonces $s_{n+1} = s_n - a_n x_n$.

respectivos equipos —llamados 1, 2 y 3— fracasen, es de 0.40, 0.60 y 0.80. En consecuencia, la probabilidad actual de que los tres equipos fracasen es (0.40)(0.60)(0.80) = 0.192. Como el objetivo es minimizar la probabilidad de fracaso, se asignan al proyecto dos científicos de más alto nivel.

En la tabla 10.2 se proporciona la probabilidad estimada de que los equipos respectivos fracasen si se les asigna 0, 1 o 2 científicos para que colaboren con ellos. Sólo se consideran números enteros de científicos puesto que cada uno de ellos debe dedicar su atención completa a un equipo. El problema es determinar cómo deben asignarse los dos científicos adicionales para minimizar la probabilidad de que los tres equipos fracasen.

Formulación. En razón de que tanto el ejemplo 2 como el 3 son problemas de distribución de esfuerzo, su estructura básica es muy semejante. En este caso, los científicos sustituyen a las brigadas médicas como el recurso con que se cuenta y los equipos de investigación sustituyen a los países que juegan el papel de actividades. Entonces, en lugar de brigadas médicas que se destinan a países, se asignarán científicos a equipos de investigación. La única diferencia básica entre los dos problemas está en su función objetivo.

Debido a que el número de científicos y equipos es tan pequeño, este problema se puede resolver con facilidad mediante un proceso de enumeración exhaustiva. El procedimiento de solución por programación dinámica se presenta con propósitos explicativos.

En este caso, la etapa n ($n = 1, 2, 3$) corresponde al equipo de investigación n, mientras que el estado s_n es el número de científicos *todavía disponible* que debe asignarse a los equipos restantes. Las variables de decisión x_n($n = 1, 2, 3$) son el número de científicos adicionales que se asigna al equipo n.

Sea $p_i(x_i)$ la probabilidad de fracaso del equipo i si se le asignan x_i científicos adicionales, como se ve en la tabla 10.2. Si el signo Π denota multiplicación, el objetivo del gobierno será elegir x_1, x_2, x_3 para

$$\text{Minimizar } \prod_{i=1}^{3} p_i(x_i) = p_1(x_1)p_2(x_2)p_3(x_3),$$

sujeta a

$$\sum_{i=1}^{3} x_i = 2$$

y

las x_i son enteros no negativos.

En consecuencia, $f_n(s_n, x_n)$ en este problema es

$$f_n(s_n, x_n) = p_n(x_n) \cdot \text{mín} \prod_{i=n+1}^{3} p_i(x_i),$$

donde el mínimo se toma sobre $x_{n+1}, \ldots, x_3$, tal que

$$\sum_{i=n}^{3} x_i = s_n$$

■ **TABLA 10.2** Datos del problema del proyecto espacial del gobierno

	Probabilidades de fracaso		
	Equipo		
Nuevos científicos	**1**	**2**	**3**
0	0.40	0.60	0.80
1	0.20	0.40	0.50
2	0.15	0.20	0.30

y

las x_i son enteros no negativos,

para $n = 1, 2, 3$. Así,

$$f_n^*(s_n) = \min_{x_n=0,1,\ldots,s_n} f_n(s_n, x_n),$$

donde

$$f_n(s_n, x_n) = p_n(x_n) \cdot f_{n+1}^*(s_n - x_n)$$

(con f_4^* igual a 1). En la figura 10.7 se resumen estas relaciones básicas.

Por tanto, en este caso, la *relación recursiva* entre las funciones f_1^*, f_2^* y f_3^* es

$$f_n^*(s_n) = \min_{x_n=0,1,\ldots,s_n} \{p_n(x_n) \cdot f_{n+1}^*(s_n - x_n)\}, \qquad \text{para } n = 1, 2,$$

y cuando $n = 3$,

$$f_3^*(s_3) = \min_{x_3 = 0,1,\ldots,s_3} p_3(x_3).$$

Procedimientos de solución. Los cálculos de programación dinámica que se obtienen son los siguientes:

$n = 3$:

s_3	$f_3^*(s_3)$	x_3^*
0	0.80	0
1	0.50	1
2	0.30	2

$n = 2$:

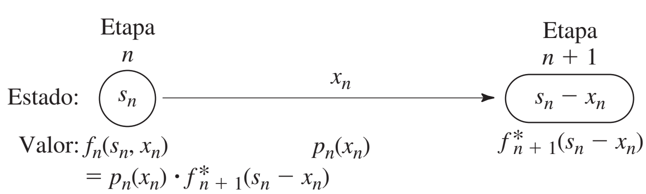
s_2 \ x_2	$f_2(s_2, x_2) = p_2(x_2) \cdot f_3^*(s_2 - x_2)$			$f_2^*(s_2)$	x_2^*
	0	1	2		
0	0.48			0.48	0
1	0.30	0.32		0.30	0
2	0.18	0.20	0.16	0.16	2

$n = 1$:

s_1 \ x_1	$f_1(s_1, x_1) = p_1(x_1) \cdot f_2^*(s_1 - x_1)$			$f_1^*(s_1)$	x_1^*
	0	1	2		
2	0.064	0.060	0.072	0.060	1

Por tanto, la solución óptima debe tener $x_1^* = 1$, lo que hace $s_2 = 2 - 1 = 1$, y con esto $x_2^* = 0$, con lo que $s_3 = 1 - 0 = 1$ y $x_3^* = 1$. De esta manera, los equipos 1 y 3 deben recibir un científico adicional cada uno. La nueva probabilidad de que los tres equipos fracasen es de 0.060.

■ **FIGURA 10.7**
Estructura básica del problema del proyecto espacial del gobierno.



Hasta ahora, todos los ejemplos han presentado variables de estado s_n *discretas* en cada etapa. Aún más, todos han sido *reversibles* en el sentido de que, de hecho, el procedimiento de solución se puede mover hacia atrás *o bien* hacia adelante etapa por etapa. (Esta última alternativa se reduce a reenumerar las etapas en el orden inverso y después aplicar el procedimiento estándar.) Esta propiedad de reversibilidad es una característica general de los problemas de distribución de esfuerzo, como los ejemplos 2 y 3, en razón de que las actividades (etapas) se pueden ordenar de cualquier manera.

El siguiente ejemplo es diferente en ambos aspectos; en lugar de estar restringida a valores enteros, la variable de estado s_n en la etapa n es una variable *continua* y puede tomar *cualquier* valor dentro de ciertos intervalos. Como s_n ahora tiene un número infinito de valores, ya no es posible considerar cada uno de sus valores posibles en forma individual. Ahora la solución para $f_n^*(s_n)$ y el valor x_n^* deben expresarse como *funciones* de s_n. Aún más, este ejemplo *no* es reversible porque sus etapas corresponden a *periodos temporales*, por lo que el procedimiento de solución debe ir hacia atrás.

Antes de proceder de manera directa con el siguiente ejemplo, al lector le puede resultar útil revisar los **dos ejemplos adicionales** de programación dinámica determinística que se presentan en la sección Worked Examples del sitio en internet de este libro. El primero se refiere a la planeación de producción e inventarios en cierto número de periodos temporales. Como en los ejemplos que se han estudiado hasta ahora, tanto la variable de estado como la variable de decisión en cada etapa son discretas. Sin embargo, este ejemplo no es reversible porque las etapas corresponden a periodos en el tiempo. Tampoco es un problema de distribución de esfuerzo. El segundo ejemplo es un problema de programación no lineal con dos variables y una sola restricción. Por tanto, aunque es reversible, sus variables de estado y de decisión son continuas. Sin embargo, a diferencia del ejemplo siguiente —que tiene cuatro variables continuas y, por ende, cuatro etapas—, éste tiene sólo dos etapas, por lo que se puede resolver relativamente rápido con programación dinámica y un poco de cálculo.

EJEMPLO 4 Programación del nivel de empleados

La carga de trabajo del LOCAL JOB SHOP está sujeta a grandes fluctuaciones que dependen de la temporada. Sin embargo, es difícil contratar y costoso capacitar a los operadores de las máquinas, por lo que el administrador rechaza la idea de despedir trabajadores durante las temporadas bajas. Tampoco quiere mantener su nómina de temporadas altas cuando no se requiere. Más aún, en forma definitiva se opone a pagar tiempo extra en forma regular. Como todos los trabajos se hacen sobre pedido, no es posible acumular un inventario durante las temporadas bajas. Por tanto, la administración tiene que determinar cuál debe ser la política sobre los niveles de empleados.

A continuación se proporcionan las estimaciones sobre la mano de obra que se requerirá durante las cuatro temporadas del año en un futuro cercano:

Temporada	Primavera	Verano	Otoño	Invierno	Primavera
Requerimientos	255	220	240	200	255

No se permitirá que el nivel de empleados baje de estos niveles. Cualquier contratación más alta se desperdicia con un costo aproximado de $2 000 por persona por temporada. Se estima que los costos de contratación y despido son tales que el costo de cambiar el nivel de empleados de una temporada a la siguiente es igual a $200 multiplicado por el cuadrado de la diferencia de nivel. Es posible contar con niveles fraccionales gracias a que hay algunos empleados de tiempo parcial, y los datos de costos se aplican igual a estas fracciones.

Formulación. Si el análisis se basa en los datos disponibles, puede observarse que no vale la pena que ningún nivel de empleo sea mayor que 255, que corresponde a los requerimientos de la temporada pico. De esta manera, en primavera el número de empleados deberá ser de 255 y el problema se reduce a encontrar el nivel de las otras tres temporadas.

Para la formulación de programación dinámica, las temporadas deben ser las etapas. En realidad existe un número indefinido de etapas pues el problema se proyecta hacia un futuro indefinido.

Sin embargo, cada año comienza un ciclo idéntico y, como se conoce el nivel de empleados de la primavera, es factible tomar en cuenta sólo un ciclo de cuatro temporadas que termine en primavera, como se resume en seguida.

Etapa 1 = verano,
Etapa 2 = otoño,
Etapa 3 = invierno,
Etapa 4 = primavera.
x_n = nivel de empleados de la etapa n ($n = 1, 2, 3, 4$).
($x_4 = 255$.)

Es necesario que la temporada de primavera sea la última etapa porque debe conocerse o poderse obtener el valor óptimo de la variable de decisión de cada estado en esta última etapa, sin tener que considerar otras. En todas las demás etapas, la solución del nivel óptimo de empleo debe tomar en cuenta el efecto sobre los costos de la siguiente temporada.

Sea

r_n = mano de obra mínima que se requiere en la etapa n,

donde estos requerimientos se dieron antes como $r_1 = 220$, $r_2 = 240$, $r_3 = 200$ y $r_4 = 255$. Así, los únicos valores factibles de x_n son

$$r_n \leq x_n \leq 255.$$

Si se hace referencia a los datos de costo que se proporcionaron en el enunciado del problema, se tiene

$$\text{Costo en la etapa } n = 200(x_n - x_{n-1})^2 + 2\,000(x_n - r_n).$$

Observe que el costo de la etapa actual depende sólo de la decisión x_n que se tome en esa etapa y del número de empleados de la etapa anterior x_{n-1}. Así, el nivel de empleados anterior es toda la información sobre el estado actual que se necesita para determinar la política óptima de ahí en adelante. Por ello, el estado s_n de la etapa n es

$$\text{Estado } s_n = x_{n-1}.$$

Cuando $n = 1$, $s_1 = x_0 = x_4 = 255$.

Para facilitar la referencia cuando se trabaje en el problema, en la tabla 10.3 se proporciona un resumen de los datos anteriores.

El objetivo en este problema es elegir a x_1, x_2, x_3 (con $x_0 = x_4 = 255$) para

$$\text{Minimizar} \quad \sum_{i=1}^{4} [200(x_i - x_{i-1})^2 + 2\,000(x_i - r_i)],$$

sujeta a

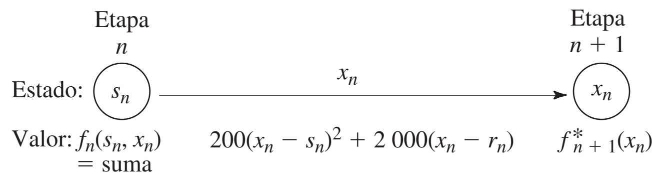
$$r_i \leq x_i \leq 255, \qquad \text{para } i = 1, 2, 3, 4.$$

Entonces, de la etapa n en adelante ($n = 1, 2, 3, 4$), debido a que $s_n = x_{n-1}$

$$f_n(s_n, x_n) = 200(x_n - s_n)^2 + 2\,000(x_n - r_n) + \min_{r_i \leq x_i \leq 255} \sum_{i=n+1}^{4} [200(x_i - x_{i-1})^2 + 2\,000(x_i - r_i)],$$

TABLA 10.3 Datos del problema del local de Job Shop

n	r_n	Factible x_n	Posible $s_n = x_{n-1}$	Costo
1	220	$220 \leq x_1 \leq 255$	$s_1 = 255$	$200(x_1 - 255)^2 + 2\,000(x_1 - 220)$
2	240	$240 \leq x_2 \leq 255$	$220 \leq s_2 \leq 255$	$200(x_2 - x_1)^2 + 2\,000(x_2 - 240)$
3	200	$200 \leq x_3 \leq 255$	$240 \leq s_3 \leq 255$	$200(x_3 - x_2)^2 + 2\,000(x_3 - 200)$
4	255	$x_4 = 255$	$200 \leq s_4 \leq 255$	$200(255 - x_3)^2$



■ **FIGURA 10.8**
Estructura básica del problema del Local Job Shop.

donde esta sumatoria es igual a cero cuando $n = 4$ (porque no tiene términos). También,

$$f^*_n(s_n) = \min_{r_n \le x_n \le 255} f_n(s_n, x_n).$$

Entonces,

$$f_n(s_n, x_n) = 200(x_n - s_n)^2 + 2\,000(x_n - r_n) + f^*_{n+1}(x_n)$$

con f^*_5 definida como cero, pues los costos después de la etapa 4 son irrelevantes para el análisis. En la figura 10.8 se presenta un resumen de estas relaciones básicas.

En consecuencia, la relación recursiva entre las funciones f^*_n es

$$f^*_n(s_n) = \min_{r_n \le x_n \le 255} \{200(x_n - s_n)^2 + 2\,000(x_n - r_n) + f^*_{n+1}(x_n)\}.$$

La programación dinámica utiliza esta relación para identificar en forma sucesiva estas funciones —$f^*_4(s_4), f^*_3(s_3), f^*_2(s_2), f^*_1(255)$— y la x_n correspondiente que minimiza.

Procedimiento de solución. *Etapa 4:* Si se inicia en la última etapa ($n = 4$), se sabe que $x^*_4 = 255$, de manera que los resultados son:

$n = 4$:

s_4	$f^*_4(s_4)$	x^*_4
$200 \le s_4 \le 255$	$200(255 - s_4)^2$	255

Etapa 3: En el caso del problema que consiste nada más en las *dos* últimas etapas ($n = 3$), la relación recursiva se reduce a

$$\begin{aligned} f^*_3(s_3) &= \min_{200 \le x_3 \le 255} \{200(x_3 - s_3)^2 + 2\,000(x_3 - 200) + f^*_4(x_3)\} \\ &= \min_{200 \le x_3 \le 255} \{200(x_3 - s_3)^2 + 2\,000(x_3 - 200) + 200(255 - x_3)^2\}, \end{aligned}$$

donde los valores posibles de s_3 son $240 \le s_3 \le 255$.

Una manera de obtener el valor de x_3 que minimiza $f_3(s_3, x_3)$ para cualquier valor dado de s_3 es el método gráfico que se muestra en la figura 10.9.

Sin embargo, una forma más rápida es usar *cálculo*. Se quiere obtener el valor de x_3 que minimiza, en términos de s_3 donde s_3 tiene un valor fijo (aunque desconocido). Se iguala a cero la primera derivada (parcial) de $f_3(s_3, x_3)$ con respecto a x_3,

$$\begin{aligned} \frac{\partial}{\partial x_3} f_3(s_3, x_3) &= 400(x_3 - s_3) + 2\,000 - 400(255 - x_3) \\ &= 400(2x_3 - s_3 - 250) \\ &= 0, \end{aligned}$$

lo que conduce a

$$x^*_3 = \frac{s_3 + 250}{2}.$$

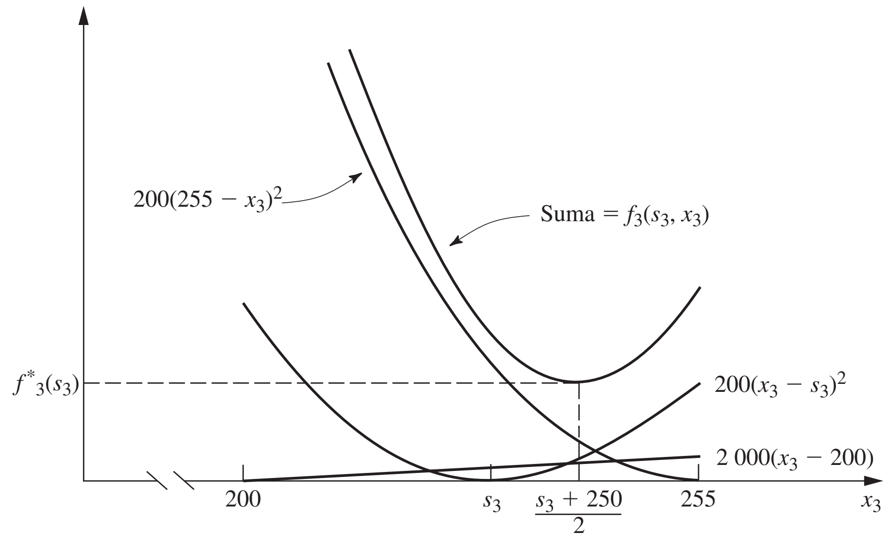


FIGURA 10.9
Solución gráfica de $f_3^*(s_3)$ del problema del Local Job Shop.

Como la segunda derivada es positiva y como esta solución se encuentra en el intervalo factible de x_3 ($200 \le x_3 \le 255$), para todos los valores posibles de s_3 ($240 \le s_3 \le 255$), sin duda se trata del mínimo que se busca.

Observe una diferencia clave entre la naturaleza de esta solución y la que se obtuvo en los ejemplos anteriores en donde sólo había que considerar unos cuantos estados posibles. Ahora se tiene un número *infinito* de estados posibles ($240 \le s_3 \le 255$), por lo que ya no es factible obtener por separado el valor de x_3^* para cada valor de s_3. En lugar de ello, se obtiene x_3^* como una *función* de la incógnita s_3.

Al utilizar

$$f_3^*(s_3) = f_3(s_3, x_3^*) = 200\left(\frac{s_3 + 250}{2} - s_3\right)^2 + 200\left(255 - \frac{s_3 + 250}{2}\right)^2 + 2\,000\left(\frac{s_3 + 250}{2} - 200\right)$$

y reducir de manera algebraica esta expresión, se obtienen los resultados que se requieren para el problema de la tercera etapa, que se resumen de la siguiente manera:

$n = 3$:

s_3	$f_3^*(s_3)$	x_3^*
$240 \le s_3 \le 255$	$50(250 - s_3)^2 + 50(260 - s_3)^2 + 1\,000(s_3 - 150)$	$\frac{s_3 + 250}{2}$

Etapa 2: Los problemas de la segunda etapa ($n = 2$) y la primera etapa ($n = 1$) se resuelven en forma parecida. Para $n = 2$,

$$\begin{aligned} f_2(s_2, x_2) &= 200(x_2 - s_2)^2 + 2\,000(x_2 - r_2) + f_3^*(x_2) \\ &= 200(x_2 - s_2)^2 + 2\,000(x_2 - 240) \\ &\quad + 50(250 - x_2)^2 + 50(260 - x_2)^2 + 1\,000(x_2 - 150). \end{aligned}$$

Los valores posibles de s_2 son $220 \le s_2 \le 255$ y la región factible de x_2 es $240 \le x_2 \le 255$. El problema es minimizar el valor de x_2 en este intervalo, de modo que

$$f_2^*(s_2) = \min_{240 \le x_2 \le 255} f_2(s_2, x_2).$$

Si se iguala a cero la derivada parcial respecto de x_2:

$$\frac{\partial}{\partial x_2} f_2(s_2, x_2) = 400(x_2 - s_2) + 2\,000 - 100(250 - x_2) - 100(260 - x_2) + 1\,000$$
$$= 200(3x_2 - 2s_2 - 240)$$
$$= 0$$

se obtiene

$$x_2 = \frac{2s_2 + 240}{3}.$$

Puesto que

$$\frac{\partial^2}{\partial x_2^2} f_2(s_2, x_2) = 600 > 0,$$

este valor de x_2 sería el valor que da el mínimo *si fuera factible* ($240 \le x_2 \le 255$). En el caso de los valores posibles de s_2 ($220 \le s_2 \le 255$), de hecho esta solución es factible sólo si $240 \le s_2 \le 255$.

En consecuencia, todavía es necesario obtener el valor factible de x_2 que minimiza $f_2(s_2, x_2)$ cuando $220 \le s_2 < 240$. La clave para analizar el comportamiento de $f_2(s_2, x_2)$ dentro de la región factible de x_2 es nuevamente la derivada parcial de $f_2(s_2, x_2)$. Cuando $s_2 < 240$,

$$\frac{\partial}{\partial x_2} f_2(s_2, x_2) > 0, \qquad \text{para } 240 \le x_2 \le 255,$$

de manera que $x_2 = 240$ es el valor deseado que minimiza.

El siguiente paso es sustituir estos valores de x_2 en $f_2(s_2, x_2)$ para obtener $f_2^*(s_2)$ para $s_2 \ge 240$ y $s_2 < 240$. Con esto se obtiene

$n = 2$:

s_2	$f_2^*(s_2)$	x_2^*
$220 \le s_2 \le 240$	$200(240 - s_2)^2 + 115\,000$	240
$240 \le s_2 \le 255$	$\frac{200}{9}[(240 - s_2)^2 + (255 - s_2)^2 + (270 - s_2)^2] + 2\,000(s_2 - 195)$	$\frac{2s_2 + 240}{3}$

Etapa 1: En el problema de la primera etapa ($n = 1$),

$$f_1(s_1, x_1) = 200(x_1 - s_1)^2 + 2\,000(x_1 - r_1) + f_2^*(x_1).$$

Como $r_1 = 220$, la región factible para x_1 es $220 \le x_1 \le 255$. La expresión para $f_2^*(x_1)$ es diferente en los dos intervalos $220 \le x_1 \le 240$ y $240 \le x_1 \le 255$ de esta región. Por tanto,

$$f_1(s_1, x_1) = \begin{cases} 200(x_1 - s_1)^2 + 2\,000(x_1 - 220) + 200(240 - x_1)^2 + 115\,000, & \text{si } 220 \le x_1 \le 240 \\ 200(x_1 - s_1)^2 + 2\,000(x_1 - 220) + \frac{200}{9}[(240 - x_1)^2 + (255 - x_1)^2 + (270 - x_1)^2] \\ \quad + 2\,000(x_1 - 195), & \text{si } 240 \le x_1 \le 255. \end{cases}$$

Si se considera primero el caso en donde $220 \le x_1 \le 240$ se tiene

$$\frac{\partial}{\partial x_1} f_1(s_1, x_1) = 400(x_1 - s_1) + 2\,000 - 400(240 - x_1)$$
$$= 400(2x_1 - s_1 - 235).$$

Se sabe que $s_1 = 255$ (nivel de empleados en primavera), por lo que

$$\frac{\partial}{\partial x_1} f_1(s_1, x_1) = 800(x_1 - 245) < 0$$

para toda $x_1 \leq 240$. Entonces, $x_1 = 240$ es el valor que minimiza a $f_1(s_1, x_1)$ en la región $220 \leq x_1 \leq 240$.

Cuando $240 \leq x_1 \leq 255$,

$$\begin{aligned}\frac{\partial}{\partial x_1} f_1(s_1, x_1) &= 400(x_1 - s_1) + 2\,000 \\ &\quad - \frac{400}{9}[(240 - x_1) + (255 - x_1) + (270 - x_1)] + 2\,000 \\ &= \frac{400}{3}(4x_1 - 3s_1 - 225).\end{aligned}$$

Como

$$\frac{\partial^2}{\partial x_1^2} f_1(s_1, x_1) > 0 \qquad \text{para toda } x_1,$$

se establece

$$\frac{\partial}{\partial x_1} f_1(s_1, x_1) = 0,$$

lo que conduce a

$$x_1 = \frac{3s_1 + 225}{4}.$$

Como $s_1 = 255$, se concluye que $x_1 = 247.5$ minimiza el valor de $f_1(s_1, x_1)$ en la región $240 \leq x_1 \leq 255$.

Observe que esta región ($240 \leq x_1 \leq 255$) incluye $x_1 = 240$, de manera que $f_1(s_1, 240) > f_1(s_1, 247.5)$. En el penúltimo párrafo se encontró que $x_1 = 240$ minimiza $f_1(s_1, x_1)$ en la región $220 \leq x_1 \leq 240$. En consecuencia, se puede concluir que $x_1 = 247.5$ también minimiza $f_1(s_1, x_1)$ en *toda* la región factible $220 \leq x_1 \leq 255$.

El último cálculo se hace para encontrar $f_1^*(s_1)$ para $s_1 = 255$ al sustituir $x_1 = 247.5$ en la expresión $f_1(255, x_1)$ que se cumple para $240 \leq x_1 \leq 255$. Entonces,

$$\begin{aligned}f_1^*(255) &= 200(247.5 - 255)^2 + 2\,000(247.5 - 220) \\ &\quad + \frac{200}{9}[2(250 - 247.5)^2 + (265 - 247.5)^2 + 30(742.5 - 575)] \\ &= 185\,000.\end{aligned}$$

Estos resultados se resumen de la siguiente manera:

$n = 1$:

s_1	$f_1^*(s_1)$	x_1^*
255	185 000	247.5

Por tanto, al regresar a través de las tablas que se obtuvieron para $n = 2$, $n = 3$ y $n = 4$, respectivamente, y establecer cada vez $s_n = x_{n-1}^*$, la solución óptima que se obtiene es $x_1^* = 247.5$, $x_2^* = 245$, $x_3^* = 247.5$, $x_4^* = 255$, con un costo estimado total por ciclo de \$185 000.

Para concluir la ejemplificación de programación dinámica determinística se analizará un problema que requiere *más de una* variable para describir el estado en cada etapa.

EJEMPLO 5 Problema de la Wyndor Glass Company

Considere el siguiente problema de programación lineal:

$$\text{Maximizar} \qquad Z = 3x_1 + 5x_2,$$

sujeta a

$$\begin{aligned} x_1 \qquad &\leq 4 \\ 2x_2 &\leq 12 \\ 3x_1 + 2x_2 &\leq 18 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

(Puede reconocerse que este modelo es el que se estableció para manejar el problema de la Wyndor Glass Co., en la sección 3.1.) Una manera de resolver problemas pequeños de programación lineal —o no lineal— como éste es emplear programación dinámica, aspecto que se ilustra a continuación.

Formulación. Este problema requiere tomar dos decisiones interrelacionadas, es decir, el nivel de la actividad 1, denotado por x_1, y el nivel de la actividad 2, denotado por x_2. Por ello, estas dos actividades se pueden interpretar como las dos etapas de una formulación de programación dinámica. Aunque es posible tomarlas en cualquier orden, sea la etapa n = la actividad n ($n = 1, 2$). Así, x_n es la variable de decisión en la etapa n.

¿Cuáles son los estados? Es decir, dado que se toma una decisión en las etapas anteriores (si las hay), ¿qué información se necesita conocer sobre el estado actual de las cosas antes de poder tomar una decisión en la etapa n? Un poco de reflexión puede sugerir que la información requerida es la *cantidad de holgura* que queda en las restricciones funcionales. El lado derecho de estas restricciones (4, 12 y 18) se interpreta como la cantidad total disponible de los recursos 1, 2 y 3, respectivamente (como se describió en la sección 3.1). Entonces el estado s_n se puede definir como

Estado s_n = cantidad de los respectivos recursos todavía disponibles para ser asignados a las actividades restantes.

(Observe que esta definición de estado es análoga a la de los problemas de distribución de esfuerzo que incluyen los ejemplos 2 y 3, excepto que ahora se tienen tres tipos de recursos en lugar de uno.) Entonces,

$$s_n = (R_1, R_2, R_3),$$

donde R_i es la cantidad disponible del recurso i ($i = 1, 2, 3$). Por tanto,

$$\begin{aligned} s_1 &= (4, 12, 18), \\ s_2 &= (4 - x_1, 12, 18 - 3x_1). \end{aligned}$$

Sin embargo, al comenzar este problema en la etapa 2, todavía no se conoce el valor de x_1, por lo que en ese momento se usa $s_2 = (R_1, R_2, R_3)$.

Al contrario de los ejemplos anteriores, este problema tiene *tres* variables de estado —es decir, un *vector de estado* con tres componentes— en lugar de una en cada etapa. Desde un punto de vista teórico, esta diferencia no es seria. Sólo significa que, en lugar de considerar todos los valores posibles de una variable de estado, deben considerarse todas las *combinaciones* posibles de valores de varias variables de estado. Sin embargo, desde un punto de vista de eficiencia de los cálculos, esta diferencia tiende a ser una complicación muy seria. Como en general el número de combinaciones puede ser tan grande como el *producto* del número de valores posibles de las respectivas variables, la cantidad de cálculos que se requieren tiende a "explotar" con rapidez cuando se introduce variables de estado adicionales. Este fenómeno ha recibido el adecuado nombre de la **calamidad de la dimensión**.

Cada una de las tres variables de estado es *continua,* por lo que, en lugar de considerar por separado las combinaciones posibles de valores, debe emplearse el método que se introdujo en el ejemplo 4 para obtener la información requerida como una *función* del estado del sistema.

Pese a estas complicaciones, el problema es lo suficientemente pequeño como para poder resolverlo sin mucha dificultad. Para hacerlo es necesario introducir la notación usual de programación dinámica. Así,

$$\begin{aligned} f_2(R_1, R_2, R_3, x_2) &= \text{contribución de la actividad 2 a } Z \text{ si el sistema se encuentra en el} \\ &\quad \text{estado } (R_1, R_2, R_3) \text{ al iniciar la etapa 2 y la decisión es } x_2 \\ &= 5x_2, \end{aligned}$$

$$\begin{aligned} f_1(4, 12, 18, x_1) &= \text{contribución de las actividades 1 y 2 a } Z \text{ si el sistema se encuentra} \\ &\quad \text{en el estado (4, 12, 18) al iniciar la etapa 1, la decisión inmediata} \\ &\quad \text{es } x_1 \text{ y después se toma una decisión óptima en la etapa 2.} \\ &= 3x_1 + \max_{\substack{2x_2 \le 12 \\ 2x_2 \le 18 - 3x_1 \\ x_2 \ge 0}} \{5x_2\}. \end{aligned}$$

De manera similar, para $n = 1, 2$,

$$f_n^*(R_1, R_2, R_3) = \max_{x_n} \; f_n(R_1, R_2, R_3, x_n),$$

donde este máximo se toma sobre los valores factibles de x_n, por lo cual, al usar la parte relevante de las restricciones del problema, se obtiene

$$(1) \quad f_2^*(R_1, R_2, R_3) = \max_{\substack{2x_2 \le R_2 \\ 2x_2 \le R_3 \\ x_2 \ge 0}} \{5x_2\},$$

$$(2) \quad f_1(4, 12, 18, x_1) = 3x_1 + f_2^*(4 - x_1, 12, 18 - 3x_1),$$

$$(3) \quad f_1^*(4, 12, 18) = \max_{\substack{x_1 \le 4 \\ 3x_1 \le 18 \\ x_1 \ge 0}} \{3x_1 + f_2^*(4 - x_1, 12, 18 - 3x_1)\}.$$

La ecuación (1) se usará para resolver el problema de 2 etapas. La ecuación (2) muestra la estructura básica de programación dinámica para este problema, como también se puede ver en la figura 10.10. La ecuación (3) proporciona una *relación recursiva* entre f_1^* y f_2^* que se usará para resolver el problema de la etapa 1.

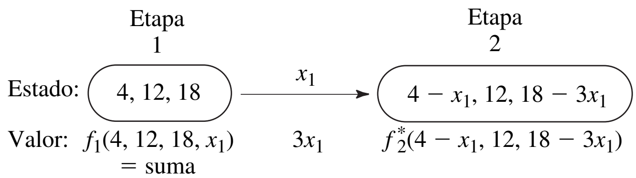
Procedimiento de solución. *Etapa 2:* Para resolver la última etapa ($n = 2$), la ecuación (1) indica que x_2^* debe ser el valor más grande de x_2 que *satisface de manera simultánea* $2x_2 \le R_2$, $2x_2 \le R_3$ y $x_2 \ge 0$. Si se supone que $R_2 \ge 0$ y $R_3 \ge 0$, para que existan soluciones factibles, este valor más grande es el más pequeño de $R_2/2$ y $R_3/2$. Por consiguiente, la solución es

$n = 2$:

(R_1, R_2, R_3)	$f_2^*(R_1, R_2, R_3)$	x_2^*
$R_2 \ge 0,\ R_3 \ge 0$	$5 \text{ mín}\left\{\frac{R_2}{2}, \frac{R_3}{2}\right\}$	$\text{mín}\left\{\frac{R_2}{2}, \frac{R_3}{2}\right\}$

Etapa 1: Para resolver el problema de dos etapas ($n = 1$), se sustituye la solución que se acaba de obtener para $f_2^*(R_1, R_2, R_3)$ en la ecuación (3). En la etapa 2,

$$(R_1, R_2, R_3) = (4 - x_1, 12, 18 - 3x_1),$$



■ **FIGURA 10.10**
Estructura básica del problema de programación lineal de la Wyndor Glass Co.

de manera que

$$f_2^*(4 - x_1, 12, 18 - 3x_1) = 5 \text{ mín} \left\{ \frac{R_2}{2}, \frac{R_3}{2} \right\} = 5 \text{ mín} \left\{ \frac{12}{2}, \frac{18 - 3x_1}{2} \right\}$$

es la solución específica que se tiene que sustituir en la ecuación (3). Después de combinar las restricciones sobre x_1, esta ecuación se convierte en

$$f_1^*(4, 12, 18) = \max_{0 \leq x_1 \leq 4} \left\{ 3x_1 + 5 \text{ mín} \left\{ \frac{12}{2}, \frac{18 - 3x_1}{2} \right\} \right\}.$$

Observe que en el intervalo factible $0 \leq x_1 \leq 4$,

$$\text{mín} \left\{ \frac{12}{2}, \frac{18 - 3x_1}{2} \right\} = \begin{cases} 6 & \text{si } 0 \leq x_1 \leq 2 \\ 9 - \frac{3}{2} x_1 & \text{si } 2 \leq x_1 \leq 4, \end{cases}$$

de manera que

$$3x_1 + 5 \text{ mín} \left\{ \frac{12}{2}, \frac{18 - 3x_1}{2} \right\} = \begin{cases} 3x_1 + 30 & \text{si } 0 \leq x_1 \leq 2 \\ 45 - \frac{9}{2} x_1 & \text{si } 2 \leq x_1 \leq 4 \end{cases}$$

Como las dos funciones

$$\max_{0 \leq x_1 \leq 2} \{3x_1 + 30\} \qquad \text{y} \qquad \max_{2 \leq x_1 \leq 4} \left\{ 45 - \frac{9}{2} x_1 \right\}$$

adquieren su máximo en $x_1 = 2$, se concluye que $x_1^* = 2$ y que este máximo es 36, como se expresa en la siguiente tabla.

$n = 1$:

(R_1, R_2, R_3)	$f_1^*(R_1, R_2, R_3)$	x_1^*
(4, 12, 18)	36	2

Como el valor $x_1^* = 2$ conduce a

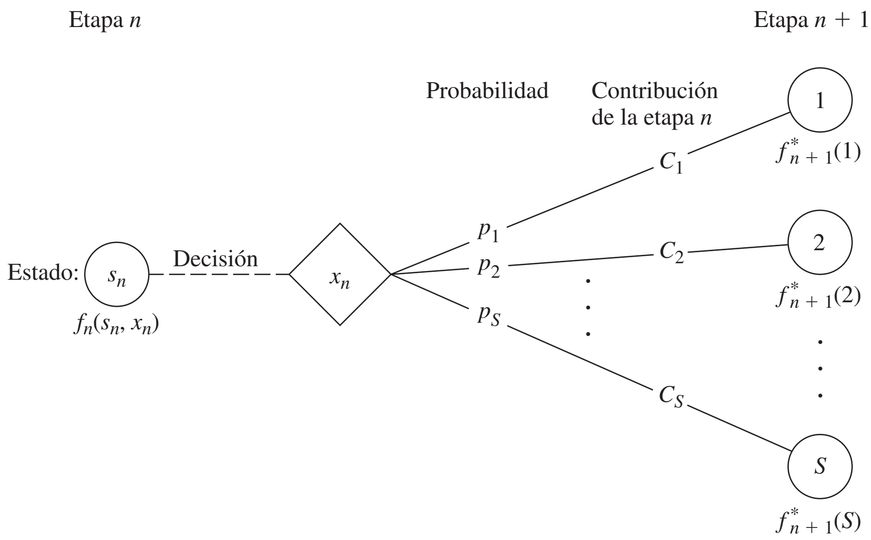
$$R_1 = 4 - 2 = 2, \qquad R_2 = 12, \qquad R_3 = 18 - 3(2) = 12$$

para la etapa 2, la tabla de $n = 2$ lleva a $x_2^* = 6$. En consecuencia, $x_1^* = 2$, $x_2^* = 6$ es la solución óptima para este problema, lo que coincide con la solución original que se encontró en la sección 3.1; la tabla de $n = 1$ muestra que el valor de Z que resulta es 36.

Hasta el momento, usted ha visto una gran variedad de aplicaciones de la programación dinámica y en la sección siguiente estudiará todavía más. Sin embargo, estos ejemplos sólo tocan el tema de manera superficial. Por ejemplo, en el capítulo 2, la referencia seleccionada 4 describe 47 tipos de problemas en los que se puede aplicar la programación dinámica. (Esta referencia también presenta una herramienta de software que puede utilizarse para resolver estos tipos de problemas.) El punto en común que aparece en estas aplicaciones de programación dinámica es la necesidad de tomar una serie de decisiones interrelacionadas y la forma eficiente que proporciona la programación dinámica para encontrar una combinación de decisiones óptima.

10.4 PROGRAMACIÓN DINÁMICA PROBABILÍSTICA

La programación dinámica *probabilística* difiere de la determinística en que el estado de la siguiente etapa *no* está determinado por completo por el estado y la política de decisión de la etapa actual. En su lugar, existe una *distribución de probabilidad* para determinar cuál será el siguiente estado. Sin embargo, esta distribución de probabilidad queda completamente determinada por el estado y



■ **FIGURA 10.11**
Estructura básica de programación dinámica probabilística.

la política de decisión de la etapa actual. En el diagrama de la figura 10.11 se describe la estructura básica que resulta para los problemas de programación dinámica probabilística.

En lo que se refiere a este diagrama, sea S el número de estados posibles en la etapa $n + 1$ y etiquete estos estados en el lado derecho con $1, 2, \ldots, S$. El sistema cambia al estado i con probabilidad p_i $(i = 1, 2, \ldots, S)$ dados el estado s_n y la decisión x_n en la etapa n. Si el sistema cambia al estado i, C_i es la contribución de la etapa n a la función objetivo.

Cuando se expande la figura 10.11 para incluir todos los estados y las decisiones posibles en todas las etapas, se obtiene lo que con frecuencia se conoce como **árbol de decisión**, el cual, si no es muy grande, proporciona una forma útil de resumir las distintas posibilidades.

Debido a la estructura probabilística, la relación entre $f_n(s_n, x_n)$ y $f^*_{n+1}(s_{n+1})$ necesariamente es más complicada que en el caso determinístico. La forma exacta de esta relación dependerá de la forma global de la función objetivo.

Para ilustrar lo anterior, suponga que el objetivo es *minimizar la suma esperada* de las contribuciones de las etapas individuales. En este caso, $f_n(s_n, x_n)$ representa la suma esperada mínima de la etapa n en adelante, *dado* que en la etapa n el estado es s_n y la política de decisión es x_n. En consecuencia,

$$f_n(s_n, x_n) = \sum_{i=1}^{S} p_i[C_i + f^*_{n+1}(i)],$$

con

$$f^*_{n+1}(i) = \min_{x_{n+1}} f_{n+1}(i, x_{n+1}),$$

donde la minimización se toma sobre todos los valores *factibles* de x_{n+1}.

El ejemplo 6 tiene esta misma forma. El ejemplo 7 ilustrará una distinta.

EJEMPLO 6 Determinación de holguras por rechazos

La HIT-AND-MISS MANUFACTURING COMPANY ha recibido un pedido para surtir un artículo de tipo especial. El cliente ha especificado estándares de calidad tan rigurosos que es posible que el fabricante tenga que producir más de un artículo para obtener uno aceptable. El número *adicional* de artículos que se producen en una corrida de producción se llama *holgura por rechazos.* Es una práctica común incluir una holgura por rechazos cuando se produce el pedido de un cliente y, en este caso, parece conveniente.

El fabricante estima que cada unidad producida de este tipo tiene una probabilidad de $\frac{1}{2}$ de ser *aceptable* y una probabilidad de $\frac{1}{2}$ de ser *defectuosa* (sin posibilidad de corrección). En consecuen-

cia, el número de unidades aceptables producidas en un lote de tamaño L tendrá una *distribución binomial*, es decir, la probabilidad de producir cero artículos aceptables en ese lote es $(\frac{1}{2})^L$.

Los costos marginales de producción se estiman en \$100 por unidad (aunque sea defectuosa), y los artículos adicionales se desperdician. Además, se incurre en costos fijos de \$300 cada vez que se pone en marcha el proceso de producción para elaborar este artículo y se requiere una preparación completa a este mismo costo para cada corrida de producción subsecuente si el procedimiento de inspección revela que en todo un lote no hubo una unidad aceptable. El fabricante tiene tiempo para realizar hasta tres corridas de producción. Si al final de la tercera corrida no obtiene un artículo aceptable, el costo ocasionado por la venta perdida y las multas será de \$1 600.

El objetivo es determinar la política sobre el tamaño del lote (1 + holgura por rechazos) de las corridas de producción que se requieren, de tal modo que se minimice el costo total esperado por el fabricante.

Formulación. La formulación de este problema de programación dinámica es:

Etapa n = n-ésima corrida de producción ($n = 1, 2, 3$),
x_n = tamaño del lote de la etapa n,
Etapa s_n = número de productos aceptables por obtener (1 o 0) al iniciar la etapa n.

Así, el estado es $s_1 = 1$ en la etapa 1. Si en lo subsecuente se obtiene al menos un artículo aceptable, el estado cambia a $s_n = 0$, después de lo cual no se incurre en costos adicionales.

Según el objetivo que se estableció para el problema,

$f_n(s_n, x_n)$ = costo total esperado en las etapas n, ..., 3 si en la etapa n el sistema comienza en el estado s_n, la decisión inmediata es x_n, y en adelante se toman decisiones óptimas,

$$f_n^*(s_n) = \min_{x_n=0,\,1,\,\ldots} f_n(s_n, x_n),$$

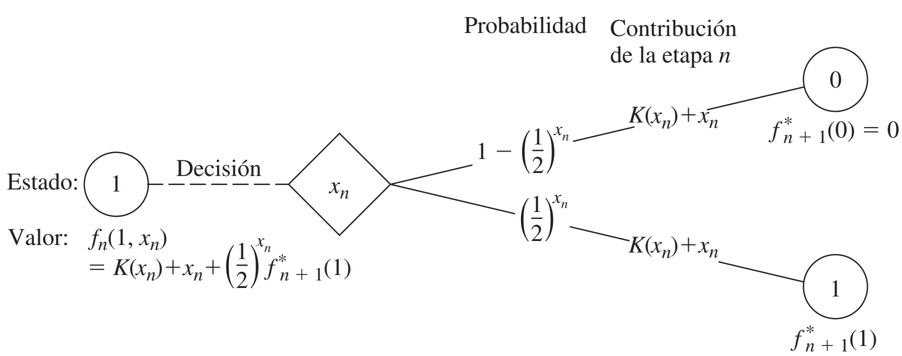
donde $f_n^*(0) = 0$. Si se usa \$100 como unidad monetaria, la contribución al costo de la etapa n es $[K(x_n) + x_n]$ independientemente del siguiente estado, donde $K(x_n)$ es una función de x_n tal que

$$K(x_n) = \begin{cases} 0, & \text{si } x_n = 0 \\ 3, & \text{si } x_n > 0. \end{cases}$$

Por tanto, para $s_n = 1$,

$$\begin{aligned} f_n(1, x_n) &= K(x_n) + x_n + \left(\frac{1}{2}\right)^{x_n} f_{n+1}^*(1) + \left[1 - \left(\frac{1}{2}\right)^{x_n}\right] f_{n+1}^*(0) \\ &= K(x_n) + x_n + \left(\frac{1}{2}\right)^{x_n} f_{n+1}^*(1) \end{aligned}$$

[donde $f_4^*(1)$ es igual a 16, el costo terminal por no haber obtenido artículos aceptables]. En la figura 10.12 se proporciona un resumen de estas relaciones básicas.



■ **FIGURA 10.12** Estructura básica del problema de programación lineal de la Hit-and-Miss Manufacturing Co.

En consecuencia, la relación recursiva de los cálculos de programación dinámica es

$$f_n^*(1) = \min_{x_n=0,\ 1,\ \ldots} \left\{K(x_n) + x_n + \left(\frac{1}{2}\right)^{x_n} f_{n+1}^*(1)\right\}$$

para $n = 1, 2, 3$.

Procedimiento de solución. Los cálculos que utilizan esta relación recursiva se resumen de la siguiente manera.

$n = 3$:

s_3 \ x_3	$f_3(1, x_3) = K(x_3) + x_3 + 16\left(\frac{1}{2}\right)^{x_3}$						$f_3^*(s_3)$	x_3^*
	0	1	2	3	4	5		
0	0						0	0
1	16	12	9	8	8	$8\frac{1}{2}$	8	3 o 4

$n = 2$:

s_2 \ x_2	$f_2(1, x_2) = K(x_2) + x_2 + \left(\frac{1}{2}\right)^{x_2} f_3^*(1)$					$f_2^*(s_2)$	x_2^*
	0	1	2	3	4		
0	0					0	0
1	8	8	7	7	$7\frac{1}{2}$	7	2 o 3

$n = 1$:

s_1 \ x_1	$f_1(1, x_1) = K(x_1) + x_1 + \left(\frac{1}{2}\right)^{x_1} f_2^*(1)$					$f_1^*(s_1)$	x_1^*
	0	1	2	3	4		
1	7	$7\frac{1}{2}$	$6\frac{3}{4}$	$6\frac{7}{8}$	$7\frac{7}{16}$	$6\frac{3}{4}$	2

Por tanto, la política óptima es producir dos artículos en la primera corrida de producción; si ninguno fuera aceptable, entonces habría que producir dos o tres artículos en la segunda corrida; si ninguno resultara aceptable, entonces se producirían tres o cuarto artículos en la tercera corrida. El costo total esperado de esta política es de $675.

EJEMPLO 7 Ganadora en Las Vegas

Una joven emprendedora experta en estadística cree haber desarrollado un sistema para ganar un popular juego en Las Vegas. Sus colegas no piensan que este sistema sea tan bueno, por lo que le apuestan que si comienza con tres fichas, ella no tendrá al menos cinco fichas después de tres jugadas. Cada jugada incluye apostar cualquier cantidad de las fichas disponibles y ganar o perder este mismo número de fichas. La joven cree que su sistema le dará una probabilidad de $\frac{2}{3}$ de ganar una jugada dada.

Si se supone que la experta en estadística está en lo correcto, se quiere usar programación dinámica para determinar su política óptima sobre cuántas fichas apostar (si apuesta) en cada una de las tres jugadas. La decisión en cada jugada deberá tomar en cuenta los resultados de las jugadas anteriores. El objetivo es maximizar la probabilidad de ganar la apuesta hecha a sus colegas.

Formulación. La formulación de programación dinámica para este problema es

Etapa n = n-ésima jugada del juego ($n = 1, 2, 3$),
x_n = número de fichas que debe apostar en la etapa n,
Etapa s_n = número de fichas disponibles para comenzar la etapa n.

Se eligió esta definición del estado del sistema porque proporciona la información necesaria sobre la situación actual para tomar una decisión óptima sobre cuántas fichas apostar en la siguiente jugada.

Como el objetivo es maximizar la probabilidad de que la joven gane la apuesta, la función objetivo que debe maximizarse en cada etapa es la probabilidad de terminar las tres jugadas con cinco fichas o más. (Observe que el valor de terminar con más de cinco fichas es el mismo que el de terminar con cinco fichas, ya que la apuesta se gana de las dos formas.) Así,

$f_n(s_n, x_n)$ = probabilidad de terminar las tres jugadas con cinco fichas o más, dado que la joven comienza la etapa n con s_n fichas, su decisión inmediata es x_n y en adelante toma decisiones óptimas,

$$f_n^*(s_n) = \max_{x_n=0,\,1,\,\ldots,\,s_n} f_n(s_n, x_n).$$

La expresión para $f_n(s_n, x_n)$ debe reflejar el hecho de que, en un momento dado, es posible acumular cinco fichas aun cuando pierda la siguiente jugada. Si pierde, el estado en la siguiente etapa será $s_n - x_n$, y la probabilidad de terminar con al menos cinco fichas será entonces $f_{n+1}^*(s_n - x_n)$. Por el contrario, si gana la siguiente jugada, el estado del sistema será $s_n + x_n$ y la probabilidad correspondiente será $f_{n+1}^*(s_n + x_n)$. Como se supone que la probabilidad de ganar una jugada dada es $\frac{2}{3}$, se concluye que

$$f_n(s_n, x_n) = \frac{1}{3} f_{n+1}^*(s_n - x_n) + \frac{2}{3} f_{n+1}^*(s_n + x_n)$$

[donde $f_4^*(s_4)$ se define como 0 para $s_4 < 5$ y 1 para $s_4 \geq 5$]. Entonces, no existe una contribución directa a la función objetivo por parte de la etapa n además del efecto que tiene estar en el siguiente estado. En la figura 10.13 se resumen estas relaciones básicas.

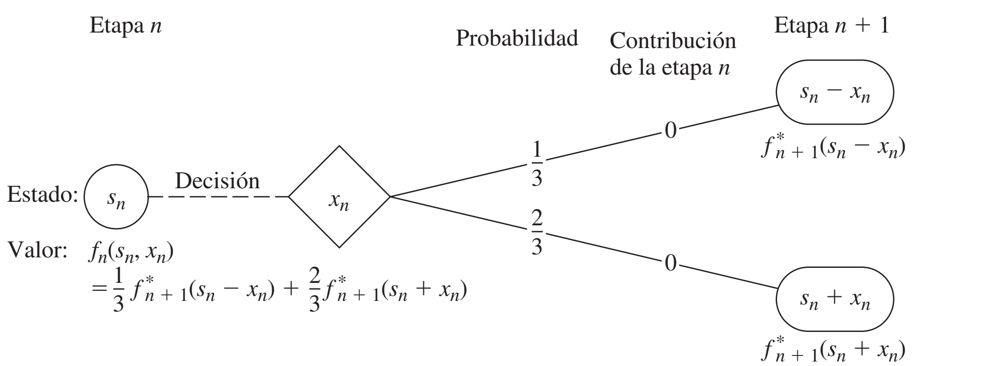
Entonces, la relación recursiva para este problema es

$$f_n^*(s_n) = \max_{x_n=0,\,1,\,\ldots,\,s_n} \left\{ \frac{1}{3} f_{n+1}^*(s_n - x_n) + \frac{2}{3} f_{n+1}^*(s_n + x_n) \right\},$$

para $n = 1, 2, 3$, con $f_4^*(s_4)$ como se definió.

Procedimiento de solución. Esta relación recursiva conduce a los siguientes resultados de cálculos.

■ **FIGURA 10.13**
Estructura básica del problema de Las Vegas.



n = 3:

s_3	$f_3^*(s_3)$	x_3^*
0	0	—
1	0	—
2	0	—
3	$\frac{2}{3}$	2 (o más)
4	$\frac{2}{3}$	1 (o más)
≥5	1	0 (o ≤ $s_3 - 5$)

n = 2:

s_2 \ x_2	$f_2(s_2, x_2) = \frac{1}{3}f_3^*(s_2 - x_2) + \frac{2}{3}f_3^*(s_2 + x_2)$					$f_2^*(s_2)$	x_2^*
	0	1	2	3	4		
0	0					0	—
1	0	0				0	—
2	0	$\frac{4}{9}$	$\frac{4}{9}$			$\frac{4}{9}$	1 o 2
3	$\frac{2}{3}$	$\frac{4}{9}$	$\frac{2}{3}$	$\frac{2}{3}$		$\frac{2}{3}$	0, 2 o 3
4	$\frac{2}{3}$	$\frac{8}{9}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{8}{9}$	1
≥5	1					1	0 (o ≤ $s_2 - 5$)

n = 1:

s_1 \ x_1	$f_1(s_1, x_1) = \frac{1}{3}f_2^*(s_1 - x_1) + \frac{2}{3}f_2^*(s_1 + x_1)$				$f_1^*(s_1)$	x_1^*
	0	1	2	3		
3	$\frac{2}{3}$	$\frac{20}{27}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{20}{27}$	1

Por tanto, la política óptima es

$$x_1^* = 1 \begin{cases} \text{si gana,} \quad x_2^* = 1 \begin{cases} \text{si gana,} & x_3^* = 0 \\ \text{si pierde,} & x_3^* = 2 \text{ o } 3. \end{cases} \\ \text{si pierde,} \quad x_2^* = 1 \text{ o } 2 \begin{cases} \text{si gana,} & x_3^* = \begin{cases} 2 \text{ o } 3 & (\text{para } x_2^* = 1) \\ 1, 2, 3 \text{ o } 4 & (\text{para } x_2^* = 2) \end{cases} \\ \text{si pierde,} & \text{la apuesta está perdida} \end{cases} \end{cases}$$

Esta política da a la experta en estadística una probabilidad de $\frac{20}{27}$ de ganar la apuesta a sus colegas.

10.5 CONCLUSIONES

La programación dinámica es una técnica muy útil para tomar una *sucesión de decisiones interrelacionadas.* Requiere la formulación de una *relación recursiva* apropiada para cada problema individual. Sin embargo, proporciona grandes ahorros computacionales en comparación con la enumeración exhaustiva para encontrar la mejor combinación de decisiones, en especial cuando se trata de problemas grandes. Por ejemplo, si un problema tiene 10 etapas con 10 estados y l0 decisiones posibles en cada etapa, la enumeración exhaustiva tendría que considerar hasta 10 mil millones de combinaciones, mientras que la programación dinámica necesita hacer como máximo mil cálculos (10 para cada estado en cada etapa).

En este capítulo se presentó sólo programación dinámica con un número *finito* de etapas. El capítulo 19 está dedicado a un tipo general de modelos de programación dinámica probabilista en donde las etapas continúan de manera indefinida, tema conocido como procesos markovianos de decisión.

REFERENCIAS SELECCIONADAS

1. Bertsekas, D. P., *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, Englewood Cliffs, NJ, 1987.
2. Denardo, E. V., *Dynamic Programming Theory and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1982.
3. Howard, R. A., "Dynamic Programming", en *Management Science*, **12**: 317-345, 1966.
4. Lew, A. y H. Mauch: *Dynamic Programming: A Computational Tool*, Springer, Nueva York, 2007.
5. Smith, D. K., *Dynamic Programming: A Practical Introduction*, Ellis Horwood, Londres, 1991.
6. Sniedovich, M., *Dynamic Programming*, Marcel Dekker, Nueva York, 1991.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN EL SITIO EN INTERNET DEL LIBRO (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos para el capítulo 10

Archivo de LINGO "Cap. 10. Dynamic Programming"

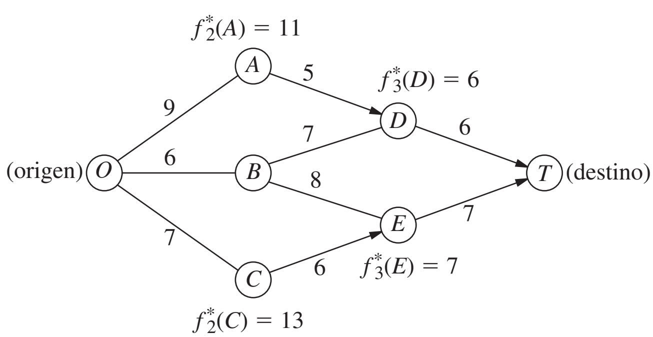
Glosario del capítulo 10

Vea el apéndice 1 para la documentación del software.

PROBLEMAS

Un asterisco en el número del problema indica que al final del libro se proporciona al menos una respuesta parcial.

10.2-1. Considere la siguiente red en la que cada número junto a una trayectoria representa la distancia real entre el par de nodos que conecta. El objetivo es encontrar la ruta más corta del origen al destino.



a) ¿Cuáles son las etapas y los estados para la formulación de programación dinámica de este problema?

b) Utilice programación dinámica para resolver el problema, pero en lugar de emplear las tablas usuales muestre el trabajo en una gráfica (similar a la figura 10.2). En particular, comience con la red dada, en la que se proporcionan los valores de $f_n^*(s_n)$ de cuatro de los nodos; después encuentre $f_2^*(B)$ y $f_1^*(O)$. Dibuje una punta de flecha para indicar la trayectoria óptima que debe tomarse al salir de estos últimos dos nodos. Por último, identifique la ruta óptima con las flechas desde el nodo *O* hasta el nodo *T*.

c) Utilice programación dinámica para resolver este problema; construya a mano las tablas usuales para $n = 3$, $n = 2$ y $n = 1$.

d) Utilice el algoritmo de la ruta más corta que se presentó en la sección 9.3 para resolver este problema. Compare este enfoque con el de los incisos *b*) y *c*).

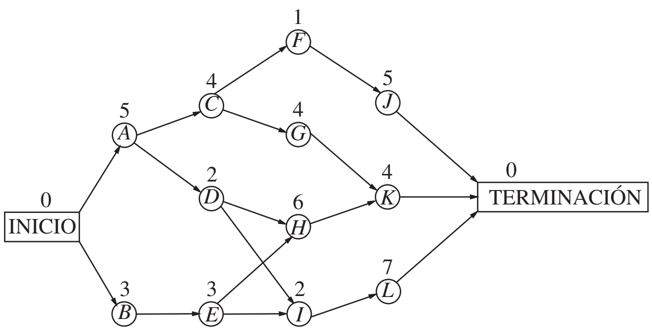
10.2-2. El gerente de ventas de una editorial de libros de texto universitarios tiene seis agentes de ventas que puede asignar a tres regiones distintas del país. Ha decidido que cada región debe tener por lo menos un agente y que cada uno de éstos debe quedar restringido a una de estas regiones, pero ahora quiere determinar cuántos agentes debe asignar a las respectivas regiones con el fin de maximizar las ventas.

La tabla de la parte superior de la siguiente columna da el incremento estimado de las ventas en cada región (en las unidades apropiadas) si se le asignan diferentes cantidades de agentes:

	Región		
Agente de ventas	**1**	**2**	**3**
1	40	24	32
2	54	47	46
3	78	63	70
4	99	78	84

a) Utilice programación dinámica para resolver este problema. En lugar de usar las tablas normales, muestre su trabajo con una gráfica de una red similar a la del problema 10.2-1. Haga lo mismo que en el problema 10.2-1*b* para obtener $f_n^*(s_n)$ de cada nodo (excepto el nodo terminal) y escriba sus valores al lado. Dibuje una punta de flecha para indicar la ruta óptima (o rutas en el caso de empates) que debe tomarse al salir de cada nodo. Por último, identifique la ruta (o rutas) óptima(s) que obtuvo a través de la red y la solución (o soluciones) óptima(s) correspondiente.
b) Utilice programación dinámica para resolver este problema; elabore las tablas normales con $n = 3$, $n = 2$ y $n = 1$.

10.2-3. Considere la siguiente red de proyecto (como se describe en la sección 9.8), donde el número sobre el nodo es el tiempo que se requiere para la actividad correspondiente. Considere el problema de encontrar la *trayectoria más larga* (el mayor tiempo total) a través de esta red desde el inicio hasta su término, puesto que la trayectoria más larga es la ruta crítica.



a) ¿Cuáles son las etapas y los estados para formular la programación dinámica de este problema?
b) Use programación dinámica para resolver este problema, pero en lugar de emplear las tablas usuales, muestre su trabajo en una gráfica. En particular, determine los valores de las distintas $f_n^*(s_n)$ junto a los nodos correspondientes y el arco óptimo que debe tomarse al salir de cada nodo mediante una punta de flecha cerca del inicio del arco. Después identifique la trayectoria óptima (la trayectoria más larga) que siguen estas puntas de flecha desde el nodo inicio hasta el de término. Si existe más de una trayectoria óptima, identifíquelas a todas.
c) Utilice programación dinámica para resolver el problema con las tablas usuales para $n = 4$, $n = 3$, $n = 2$ y $n = 1$.

10.2-4. Considere las siguientes afirmaciones sobre la solución de problemas de programación dinámica. Diga si cada una es falsa o verdadera y después justifique su respuesta con referencia a afirmaciones específicas (con la cita de la página) dentro del capítulo.
a) El procedimiento de solución utiliza una relación recursiva que permite obtener la política óptima para la etapa $(n + 1)$ dada la política óptima de la etapa n.
b) Después de completar el procedimiento de solución, si se toma por error una decisión no óptima en alguna etapa, deberá aplicarse de nuevo el procedimiento para determinar las nuevas decisiones óptimas (dada esta decisión no óptima) en las etapas subsecuentes.
c) Una vez que se ha encontrado una política óptima para todo el problema, la información necesaria para especificar la decisión óptima en una etapa en particular es el estado de esa etapa y las decisiones que se tomaron en las etapas anteriores.

10.3-1. Lea el artículo de referencia que describe en su totalidad el estudio de investigación de operaciones que se resume en el recuadro de aplicación que se presentó en la sección 10.3. Describa de manera breve la forma como se aplicó la programación dinámica en dicho estudio. Después, elabore una lista de los beneficios financieros y de otro tipo que arrojó este estudio.

10.3-2.* El propietario de una cadena de tres supermercados compró cinco cargas de fresas frescas. La distribución de probabilidad estimada de las ventas potenciales de las fresas antes de que se echen a perder difiere entre los tres supermercados. El propietario quiere saber cómo debe asignar las cinco cargas a las tiendas para maximizar la ganancia esperada.

Por razones administrativas, no quiere dividir las cargas entre las tiendas. Sin embargo, está de acuerdo en asignar cero cargas a cualquiera de ellas.

En la siguiente tabla se proporciona la ganancia estimada de cada tienda al asignar distintas cantidades de cargas:

Número de carga	Tienda 1	2	3
0	0	0	0
1	5	6	4
2	9	11	9
3	14	15	13
4	17	19	18
5	21	22	20

Utilice programación dinámica para determinar cuántas cargas debe asignarse a cada tienda para maximizar la ganancia total esperada.

10.3-3. Una estudiante universitaria cuenta con siete días para preparar los exámenes finales de cuatro cursos y quiere asignar su tiempo de estudio de la manera más eficiente posible. Necesita por lo menos un día para cada curso y quiere concentrarse sólo en un curso cada día por lo que quiere asignar uno, dos, tres o cuatro días a cada curso. Como hace poco tomó un curso de investigación de operaciones, decide aplicar programación dinámica para hacer estas asignaciones que maximicen el total de puntos obtenidos en los cuatro cursos. Estima que las distintas asignaciones en días de estudio le redituarán puntos de calificación según la siguiente tabla:

Día de estudio	Puntos de calificación estimados — Curso 1	2	3	4
1	1	5	4	4
2	3	6	6	4
3	6	8	7	5
4	8	8	9	8

Resuelva este problema con programación dinámica.

10.3-4. Una campaña política se encuentra en su última etapa y las preliminares indican que las preferencias electorales se encuentran

sumamente cerradas. Uno de los candidatos tiene suficientes fondos para comprar tiempo de TV por un total de cinco comerciales en las horas de mayor audiencia en estaciones localizadas en cuatro áreas diferentes. Con base en la información de las preliminares se hizo una estimación del número de votos adicionales que se puede ganar en las áreas de difusión según el número de comerciales que se contrate. Estas estimaciones se dan en la tabla en miles de votos:

Comerciales	Área			
	1	2	3	4
0	0	0	0	0
1	4	6	5	3
2	7	8	9	7
3	9	10	11	12
4	12	11	10	14
5	15	12	9	16

Utilice programación dinámica para determinar cómo deben distribuirse los cinco comerciales entre las cuatro áreas con el fin de maximizar el número estimado de votos ganados.

10.3-5. La presidenta de un partido político de un estado planea las próximas elecciones presidenciales. Cuenta con la colaboración de seis voluntarios para trabajar en los distritos electorales y los quiere asignar a cuatro distritos de manera que se maximice su eficacia. Ella piensa que sería ineficiente asignar un voluntario a más de un distrito, pero está dispuesta a no asignar a nadie a cualquiera de ellos si pueden lograr más en otro distrito.

En la siguiente tabla se presenta el aumento estimado del número de votos para el candidato del partido en cada distrito si se asigna distintos números de voluntarios:

Voluntarios	Distrito			
	1	2	3	4
0	0	0	0	0
1	4	7	5	6
2	9	11	10	11
3	15	16	15	14
4	18	18	18	16
5	22	20	21	17
6	24	21	22	18

Este problema tiene varias soluciones óptimas para determinar cuántos voluntarios deben asignarse a cada distrito para maximizar el incremento total estimado de la popularidad del candidato del partido. Utilice programación dinámica para encontrar todas las soluciones óptimas para que la presidenta del partido pueda hacer una selección basada en otros factores.

10.3-6. Utilice programación dinámica para resolver el problema de programación de la producción de Northern Airplane Co., que se presentó en la sección 8.1 (vea la tabla 8.7). Suponga que las cantidades producidas deben ser enteros múltiplos de cinco.

10.3-7.* Una compañía está por introducir un nuevo producto a un mercado muy competido y planea su estrategia de comercialización. Se ha tomado la decisión de introducir el producto en tres fases. La fase 1 incluye ofertas especiales de introducción a precio reducido para atraer a los compradores de primera vez. La fase 2 es una campaña intensiva de comerciales y anuncios para persuadir a estos compradores de primera vez a que continúen comprando el producto a precio normal. Se sabe que otra compañía introducirá otro nuevo producto competitivo más o menos al terminar la fase 2. En consecuencia, la fase 3 incluye una campaña de seguimiento y promoción para tratar de evitar que los clientes regulares cambien a la competencia.

Se cuenta con un presupuesto total de 4 millones de dólares para realizar esta campaña. El problema consiste en determinar cómo asignar este dinero de la manera más eficaz a las tres fases. Sea m la proporción de mercado inicial (expresada como porcentaje) que se logra en la fase 1, f_2 la fracción de este mercado que se retiene en la fase 2 y f_3 la fracción restante del porcentaje de mercado que se retiene en la fase 3. Con los datos de la siguiente tabla, aplique programación dinámica para determinar la asignación de $4 millones para maximizar el porcentaje final del mercado para el nuevo producto, es decir, maximizar mf_2f_3.

***a*)** Suponga que el dinero se debe gastar en cantidades enteras múltiplos de 1 millón de dólares en cada fase y que el mínimo permisible es 1 para la fase 1 y 0 para las fases 2 y 3. En la siguiente tabla se proporciona el efecto estimado de los gastos en cada fase:

Millones de dólares gastados	Efecto sobre el porcentaje del mercado		
	m	f_2	f_3
0	—	0.2	0.3
1	20	0.4	0.5
2	30	0.5	0.6
3	40	0.6	0.7
4	50	—	—

***b*)** Suponga que se puede gastar *cualquier* cantidad del presupuesto en cada fase, y que el efecto estimado al gastar una cantidad x_i (en unidades de *millones* de dólares) en la fase i ($i = 1, 2, 3$) es
$m = 10x_1 - x_1^2$
$f_2 = 0.40 + 0.10x_2$
$f_3 = 0.60 + 0.07x_3$.

[*Sugerencia:* Después de obtener en forma analítica las funciones $f_2^*(s)$ y $f_3^*(s)$, obtenga x_1^* de manera gráfica.]

10.3-8. Considere un sistema electrónico con cuatro componentes, cada uno de los cuales debe trabajar para que el sistema funcione. La confiabilidad del sistema se puede mejorar si se instalan varias unidades paralelas en uno o más de los componentes. La siguiente tabla muestra la probabilidad de que los respectivos componentes (indicados como componente 1, 2, 3 y 4) funcionen si constan de una, dos o tres unidades paralelas:

Unidades paralelas	Probabilidad de funcionamiento			
	Componente 1	Componente 2	Componente 3	Componente 4
1	0.5	0.6	0.7	0.5
2	0.6	0.7	0.8	0.7
3	0.8	0.8	0.9	0.9

La probabilidad de que el sistema funcione es el producto de las probabilidades de que los componentes respectivos funcionen.

En la siguiente tabla se presenta el costo (en cientos de dólares) de instalar una, dos o tres unidades paralelas en los componentes (indicados como componente 1, 2, 3 y 4) respectivos:

	Costo			
Unidades paralelas	**Componente 1**	**Componente 2**	**Componente 3**	**Componente 4**
1	1	2	1	2
2	2	4	3	3
3	3	5	4	4

Dadas las limitaciones de presupuesto, se puede gastar un máximo de 1 000 dólares.

Use programación dinámica para determinar cuántas unidades paralelas debe instalar en cada uno de los cuatro componentes para maximizar la probabilidad de que el sistema funcione.

10.3-9. Considere el siguiente problema de programación no lineal entera.

$$\text{Maximizar} \quad Z = 3x_1^2 - x_1^3 + 5x_2^2 - x_2^3,$$

sujeta a

$$x_1 + 2x_2 \leq 4$$

y

$$x_1 \geq 0, \quad x_2 \geq 0$$
$$x_1, x_2 \text{ son enteros.}$$

Use programación dinámica para resolver este problema.

10.3-10. Considere el siguiente problema de programación no lineal entera.

$$\text{Maximizar} \quad Z = 32x_1 - 2x_1^2 + 30x_2 + 20x_3,$$

sujeta a

$$3x_1 + 7x_2 + 5x_3 \leq 20$$

y

$$x_1, x_2, x_3 \text{ son enteros no negativos.}$$

Use programación dinámica para resolver el problema.

10.3-11.* Considere el siguiente problema de programación no lineal.

$$\text{Maximizar} \quad Z = 36x_1 + 9x_1^2 - 6x_1^3 + 36x_2 - 3x_2^3,$$

sujeta a

$$x_1 + x_2 \leq 3$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

Use programación dinámica para resolver el problema.

10.3-12. Resuelva el problema de programación del nivel de empleados de Local Job Shop (ejemplo 4) bajo el supuesto de que el costo total de cambiar el nivel de empleados de una temporada a la siguiente es de $100 multiplicado por el cuadrado de la diferencia de niveles.

10.3-13. Considere el siguiente problema de programación no lineal.

$$\text{Maximizar} \quad Z = 2x_1^2 + 2x_2 + 4x_3 - x_3^2$$

sujeta a

$$2x_1 + x_2 + x_3 \leq 4$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

Use programación dinámica para resolver el problema.

10.3-14. Considere el siguiente problema de programación no lineal.

$$\text{Minimizar} \quad Z = x_1^4 + 2x_2^2$$

sujeta a

$$x_1^2 + x_2^2 \geq 2.$$

(Sin restricciones de no negatividad.) Utilice programación dinámica para resolver este problema.

10.3-15. Considere el siguiente problema de programación no lineal.

$$\text{Maximizar} \quad Z = x_1^3 + 4x_2^2 + 16x_3,$$

sujeta a

$$x_1x_2x_3 = 4$$

y

$$x_1 \geq 1, \quad x_2 \geq 1, \quad x_3 \geq 1.$$

a) Utilice programación dinámica para resolver este problema cuando, además de las restricciones dadas, se requiere que todas las variables sean enteras.

b) Utilice programación dinámica para resolver el problema como se estableció (variables continuas).

10.3-16. Considere el siguiente problema de programación no lineal.

$$\text{Maximizar} \quad Z = x_1(1 - x_2)x_3,$$

sujeta a

$$x_1 - x_2 + x_3 \leq 1$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$$

Use programación dinámica para resolver el problema.

10.3-17. Considere el siguiente problema de programación lineal.

$$\text{Maximizar} \quad Z = 15x_1 + 10x_2,$$

sujeta a

$$x_1 + 2x_2 \leq 6$$
$$3x_1 + x_2 \leq 8$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

Use programación dinámica para resolver el problema.

10.3-18. Considere el siguiente problema de "costo fijo".

$$\text{Maximizar} \quad Z = 3x_1 + 7x_2 + 6f(x_3),$$

sujeta a

$$x_1 + 3x_2 + 2x_3 \leq 6$$
$$x_1 + x_2 \leq 5$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0,$$

donde

$$f(x_3) = \begin{cases} 0 & \text{si } x_3 = 0 \\ -1 + x_3 & \text{si } x_3 > 0. \end{cases}$$

Use programación dinámica para resolver el problema.

10.4-1. Un aficionado al backgamon jugará tres encuentros consecutivos con sus amigos. En cada juego tendrá la oportunidad de apostar; la cantidad apostada puede ascender a *cualquier* cantidad entre cero y la cantidad de dinero que le quede después de las apuestas en los juegos anteriores. En cada encuentro, la probabilidad de que gane el juego y, por lo tanto, la apuesta, es $\frac{1}{2}$, mientras que tiene una probabilidad de $\frac{1}{2}$ de perder y, por tanto, perder la apuesta. Comenzará con \$75 y su meta es tener \$100 al finalizar los tres juegos. (No intenta acabar con más de \$100 porque se trata de un encuentro amistoso.) El jugador quiere encontrar la política óptima de apuesta —incluidos todos los empates— que maximice la probabilidad de tener exactamente \$100 después de los tres juegos.

Utilice programación dinámica para resolver este problema.

10.4-2. Imagine que tiene \$10 000 para invertir y que tendrá la oportunidad de hacerlo en cualesquiera de dos inversiones (A o B) al principio de cada uno de los próximos tres años. Existe incertidumbre respecto del rendimiento de ambas inversiones. Si invierte en A, puede perder todo el dinero u (con probabilidad más alta) obtener \$20 000 (una ganancia de \$10 000) al final del año. Si invierte en B, puede obtener los mismos \$10 000 o (con probabilidad más baja) \$20 000 al terminar el año. Las probabilidades para que sucedan estos eventos son las siguientes:

Inversión	Cantidad obtenida (\$)	Probabilidad
A	0	0.25
	20 000	0.75
B	10 000	0.9
	20 000	0.1

Se le permite hacer (a lo sumo) *una* inversión al año y sólo puede invertir \$10 000 cada vez. (Cualquier cantidad adicional de dinero acumulada es inútil.)

a) Utilice programación dinámica para encontrar la política de inversión que maximice la cantidad de dinero que espera tener después de los tres años.

b) Utilice programación dinámica para encontrar la política de inversión que maximice la probabilidad de tener por lo menos \$20 000 después de los 3 años.

10.4-3.* Suponga que la situación del problema de la Hit-and-Miss Manufacturing Company (ejemplo 6) ha cambiado ligeramente. Después de un análisis más cuidadoso, se estima que cada artículo producido tiene una probabilidad de $\frac{2}{3}$, de ser aceptable, en lugar de $\frac{1}{2}$, de manera que la probabilidad de producir *cero* artículos aceptables en un lote de tamaño L es $(\frac{1}{3})^L$. Aún más, ahora sólo se tiene tiempo para hacer dos corridas de producción. Utilice programación dinámica para determinar la nueva política óptima para manejar este problema.

10.4-4. Reconsidere el ejemplo 7. Suponga que se cambia la apuesta a "si comienza con dos fichas, ella no tendrá cinco fichas después de cinco jugadas". Sin perder de vista los resultados que obtuvo antes, efectúe los cálculos adicionales para determinar cuál es la nueva política óptima para la joven experta en estadística.

10.4-5. El producto más importante de la Profit & Gambit Co., ha perdido dinero por un decremento de sus ventas. De hecho, durante el trimestre actual las ventas estarán 4 millones de unidades abajo del punto de equilibrio. Debido a que el rendimiento marginal por cada unidad vendida es \$5 mayor que el costo marginal, la empresa sufrirá una pérdida de \$20 millones en el trimestre. La administración debe tomar medidas expeditas para rectificar esta situación. Se tiene en consideración dos alternativas. Una consiste en abandonar el producto de inmediato, con lo que se incurriría en un costo de \$20 millones por cierre. La otra es emprender una campaña publicitaria intensiva para aumentar las ventas y después abandonar el producto (a un costo de \$20 millones) sólo si la campaña no tiene el éxito suficiente. Se formularon y analizaron planes tentativos para esta campaña que se extendería a los próximos tres trimestres —sujeta a cancelación en cualquier momento—, cuyo costo sería de \$30 millones por trimestre. Se estima que el aumento aproximado en ventas sería de 3 millones de unidades en el primer trimestre, otros 2 millones en el segundo y otro millón en el tercer trimestre. Sin embargo, por algunas variables impredecibles del mercado, existe gran incertidumbre en cuanto al efecto real que tendrá la campaña: un análisis cuidadoso indica que la estimación de cada trimestre puede estar equivocada hasta en 2 millones de unidades en cualquier dirección. (Para cuantificar la incertidumbre, suponga que los aumentos adicionales de ventas de los tres trimestres son variables aleatorias independientes con distribución uniforme entre 1 y 5 millones, entre 0 y 4 millones y entre –1 y 3 millones, respectivamente.) Si los aumentos reales son demasiado pequeños, se puede interrumpir la campaña publicitaria y discontinuar el producto al final de cualquiera de los dos próximos trimestres.

Si se emprende la campaña publicitaria intensiva y se continúa hasta terminar, se estima que las ventas seguirán en aumento por algún tiempo al mismo nivel que en el tercer trimestre (el último) de la campaña. Por tanto, si las ventas en ese trimestre todavía están por debajo del punto de equilibrio, el producto deberá dejar de producirse. De otra manera, se estima que la ganancia descontada esperada en adelante será de \$40 por cada unidad vendida que esté más arriba del punto de equilibrio en el tercer trimestre.

Utilice programación dinámica para determinar la política óptima que maximice la ganancia esperada.

CAPÍTULO 11

Programación entera

En el capítulo 3 se estudiaron muchos ejemplos de aplicaciones distintas de la programación lineal. Sin embargo, una limitación importante que impide muchas otras aplicaciones es el supuesto de divisibilidad (vea la sección 3.3), que requiere que las variables de decisión puedan tomar valores no enteros. En muchos problemas prácticos, las variables de decisión sólo tienen sentido real si su valor es entero. Por ejemplo, con frecuencia es necesario asignar a las actividades cantidades enteras de personas, máquinas o vehículos. Si el hecho de exigir valores enteros es la única diferencia que tiene un problema con la formulación de programación lineal, entonces se trata de un problema de *programación entera* (**PE**). (El nombre completo es *programación lineal entera*, pero, por lo general, el adjetivo *lineal* se omite, excepto cuando se hace una comparación con el problema más esotérico de programación no lineal entera que está fuera del alcance de este libro.)

El modelo matemático para programación entera es sencillamente el modelo de programación lineal (vea la sección 3.2) con la restricción adicional de que las variables deben tener valores enteros. Si sólo es necesario que *algunas* de las variables tengan valores enteros —y el supuesto de divisibilidad se cumple para el resto—, el modelo se conoce como de **programación entera mixta** (**PEM**). Cuando se hace la distinción entre un problema con todas las variables enteras y este caso mixto, el primero se llama de programación entera pura.

Por ejemplo, el problema de la Wyndor Glass Co., que se presentó en la sección 3.1 habría sido en realidad un problema de PE, si las dos variables de decisión x_1 y x_2 hubieran representado el número total de unidades que se debía producir de los respectivos artículos 1 y 2, en lugar de las tasas de producción. Como ambos productos —puertas de vidrio y ventanas con marco de madera— deben ser unidades completas, x_1 y x_2 deben restringirse a valores enteros.

Se realizan numerosas aplicaciones de programación entera que involucran una extensión directa de programación lineal en la que debe eliminarse el supuesto de divisibilidad. Sin embargo, existe otra área de aplicación que puede ser mucho más importante, como el problema que incluye cierto número de "decisiones tipo: sí o no". En situaciones de este tipo, las únicas dos elecciones posibles son *sí* y *no*. Por ejemplo, ¿debe emprenderse un proyecto específico? ¿Debe hacerse cierta inversión fija? ¿Debe localizarse una instalación en un sitio en particular?

Debido a que estos problemas involucran sólo dos posibilidades, este tipo de decisiones se puede representar mediante variables de decisión restringidas a sólo dos valores, por ejemplo, 0 y 1. De esta forma, la j-ésima decisión sí o no se puede representar por x_j, tal que

$$x_j = \begin{cases} 1 & \text{si la decisión } j \text{ es sí} \\ 0 & \text{si la decisión } j \text{ es no.} \end{cases}$$

Las variables de este tipo se llaman **binarias** (o variables 0-1). En consecuencia, algunas veces se hace referencia a los problemas de programación entera que contienen sólo variables binarias como problemas de **programación entera binaria** (**PEB**) (o problemas 0-1 de programación entera).

En la sección 11.1 se presenta una versión en miniatura de un problema representativo de PEB y en la sección 11.2 se estudian otras aplicaciones. Otras posibilidades de formulación con

variables binarias se analizan en la sección 11.3, mientras que en la sección 11.4 se presenta una serie de ejemplos de modelos. En las secciones 11.5 a la 11.8 se analizan las maneras de resolver los problemas de PE, que incluye los de PEB y PEM. El capítulo concluye en la sección 11.9, donde se presenta un interesante desarrollo reciente (*programación de restricciones*) que promete expandir en gran forma la capacidad de formular y resolver modelos de programación entera.

11.1 EJEMPLO PROTOTIPO

La CALIFORNIA MANUFACTURING COMPANY analiza la posibilidad de llevar a cabo una expansión mediante la construcción de una nueva fábrica ya sea en Los Ángeles o en San Francisco, o tal vez en ambas ciudades. También piensa en construir, a lo sumo, un nuevo almacén, pero la decisión sobre el lugar en donde lo instalará está restringida a la ciudad donde se construya la nueva fábrica. En la cuarta columna de la tabla 11.1 se muestra el *valor presente neto* —rendimiento total que toma en cuenta el valor del dinero en el tiempo— de cada alternativa. En la última columna se proporciona el capital que se requiere —incluido el valor presente neto— para las respectivas inversiones, donde el capital total disponible es de 10 millones de dólares. El objetivo es encontrar la combinación factible de alternativas que maximice el valor presente neto total.

El modelo PEB

Aun cuando este problema es tan pequeño que se puede resolver con rapidez por inspección —construir fábricas en ambas ciudades, pero ningún almacén—, se formulará el modelo de programación entera con propósitos ilustrativos. Todas las variables de decisión tienen la forma *binaria*

$$x_j = \begin{cases} 1 & \text{si la decisión } j \text{ es sí,} \\ 0 & \text{si la decisión } j \text{ es no,} \end{cases} \qquad (j = 1, 2, 3, 4).$$

Sea

$$Z = \text{valor presente neto de estas decisiones.}$$

Si se hace la inversión para construir una instalación dada —de manera que la variable de decisión correspondiente tenga valor de 1—, el valor presente neto estimado de estas inversiones aparece en la cuarta columna de la tabla 11.1. Si la inversión no se hace —y, por tanto, la variable de decisión es igual a 0—, el valor presente neto es 0. Entonces, con unidades de millones de dólares,

$$Z = 9x_1 + 5x_2 + 6x_3 + 4x_4.$$

La última columna de la tabla 11.1 indica que la cantidad de capital que se gastará en las cuatro instalaciones no puede exceder a 10 millones de dólares. En consecuencia, si se sigue con el uso de estas unidades de millones de dólares, una restricción del modelo es

$$6x_1 + 3x_2 + 5x_3 + 2x_4 \leq 10.$$

Como las últimas dos decisiones representan *alternativas mutuamente excluyentes* —la compañía quiere construir *cuando mucho* un almacén nuevo—, se necesita la restricción

$$x_3 + x_4 \leq 1.$$

TABLA 11.1 Datos del ejemplo de la California Manufacturing Co.

Número de decisión	Pregunta sí o no	Variable de decisión	Valor presente neto	Capital requerido
1	¿Construir la fábrica en Los Ángeles?	x_1	$9 millones	$6 millones
2	¿Construir la fábrica en San Francisco?	x_2	$5 millones	$3 millones
3	¿Construir el almacén en Los Ángeles?	x_3	$6 millones	$5 millones
4	¿Construir el almacén en San Francisco?	x_4	$4 millones	$2 millones
			Capital disponible:	$10 millones

Aún más, las decisiones 3 y 4 son *contingentes* —o condicionales— porque dependen de las decisiones 1 y 2, respectivamente (la compañía consideraría la construcción de un almacén en determinada ciudad sólo si la nueva fábrica va a estar ahí). Por tanto, en caso de tomar la decisión 3, se requiere que $x_3 = 0$ si $x_1 = 0$. Esta restricción sobre x_3 (cuando $x_1 = 0$) se impone al agregar la restricción

$$x_3 \leq x_1.$$

De manera similar, el requerimiento de que $x_4 = 0$ si $x_2 = 0$ se impone con la restricción

$$x_4 \leq x_2.$$

Por tanto, después de escribir de nuevo estas dos restricciones para que todas las variables queden en el lado izquierdo, el modelo completo de PEB es

$$\text{Maximizar} \qquad Z = 9x_1 + 5x_2 + 6x_3 + 4x_4,$$

sujeta a

$$\begin{aligned}
6x_1 + 3x_2 + 5x_3 + 2x_4 &\leq 10 \\
x_3 + x_4 &\leq 1 \\
-x_1 + x_3 &\leq 0 \\
-x_2 + x_4 &\leq 0 \\
x_j &\leq 1 \\
x_j &\geq 0
\end{aligned}$$

y

$$x_j \text{ es entera,} \qquad \text{para } j = 1, 2, 3, 4.$$

De manera equivalente, las tres últimas líneas de este modelo se pueden sustituir por una sola restricción

$$x_j \text{ es binaria,} \qquad \text{para } j = 1, 2, 3, 4.$$

Excepto por su tamaño pequeño, este ejemplo representa muchas aplicaciones reales de programación entera en las que las decisiones básicas que se toman son del tipo sí o no. Al igual que el segundo par de decisiones de este ejemplo, muchos grupos de decisiones sí o no constituyen grupos de **alternativas mutuamente excluyentes**, tales que *sólo una* decisión de ese grupo puede ser sí. Cada grupo requiere una restricción que obligue a la suma de las variables binarias correspondientes a ser igual a 1 si *exactamente* una decisión de ese grupo debe ser sí, o menor o igual a 1 (si *cuando mucho una* decisión de ese grupo puede ser sí). En ocasiones, las decisiones del tipo sí o no son **decisiones contingentes**, es decir, dependen de decisiones anteriores. Por ejemplo, se dice que una decisión es *contingente* respecto a otra si se permite que sea si *sólo si* la otra es sí. Esta situación ocurre cuando una decisión contingente implica una acción que sigue a otra y que se vuelve irrelevante, o imposible, si la otra decisión es no. La forma de la restricción que se obtiene se ilustra en la cuarta y quinta restricciones del ejemplo.

Opciones de software para resolver estos modelos

Todos los paquetes de software incluidos en el OR Courseware (Excel, LINGO/LINDO y MPL/CPLEX) contienen un algoritmo para resolver modelos de PEB (pura o mixta), así como otro para solucionar modelos generales de PE (pura o mixta) donde las variables deben ser enteras pero no binarias. Sin embargo, en razón de que las variables binarias son mucho más fáciles de manejar que las variables enteras generales, por lo común el primer algoritmo puede resolver problemas mucho más grandes que el segundo.

Cuando se utiliza Excel Solver, el procedimiento es básicamente el mismo que para programación lineal. La única diferencia surge cuando se oprime el botón de "Agregar" en el cuadro de diálogo de Solver para añadir restricciones. Además de las que se ajustan a programación lineal, se necesitan las restricciones de enteros. En el caso de variables enteras que no son binarias, éste

se logra en el cuadro de diálogo de "Agregar restricción" si se introduce el rango de variables restringidas a enteros en el lado izquierdo y después se elige "int" del menú desplegable. En el caso de variables binarias, se selecciona "bin" del menú.

En uno de los archivos de Excel para este capítulo se muestra la formulación completa en hoja de cálculo del ejemplo de la California Manufacturing Co. En la sección Worked Examples de la página en internet del libro también se incluye un **pequeño ejemplo de minimización** con dos variables restringidas a enteros. Con este caso se ilustra la formulación del modelo de PE y su solución gráfica, junto con su formulación y solución en hoja de cálculo.

Un modelo de LINGO usa la función @BIN() para especificar que las variables entre paréntesis son binarias. En el caso de una variable entera *general* (restringida a valores enteros pero no binarios), se usa la función @GIN() de la misma manera. En cualquiera de los dos casos, la función se puede anidar dentro de una instrucción @FOR para imponer estas restricciones binarias o enteras sobre un conjunto completo de variables.

En un modelo de LINDO, las restricciones binarias o enteras se insertan después de una instrucción END. Se especifica una variable X como una variable entera general con GIN X. De otra manera, para cualquier valor entero positivo de n, la instrucción GIN n especifica que las primeras n variables son enteras generales. Las variables binarias se manejan de la misma manera excepto que la palabra INTEGER se incorpora en lugar de GIN.

En el caso de un modelo de MPL, se usa la palabra clave INTEGER para designar a las variables enteras generales, mientras que BINARY indica las variables binarias. En la sección de variables de un modelo MPL, todo lo que se necesita es agregar el objetivo adecuado (INTEGER o BINARY) delante de VARIABLES para especificar que el conjunto de variables enumeradas bajo esta etiqueta es de ese tipo. Como alternativa, se puede ignorar esta especificación en la sección de variables y colocar restricciones de enteras y binarias en la sección del modelo en cualquier lugar después de las otras restricciones. En este caso, la etiqueta del conjunto de variables se convierte en sólo INTEGER o BINARY.

El complemento CPLEX de MPL incluye los últimos algoritmos para resolver modelos de PE y PEB puros o mixtos. Si elige *MIP Strategy* del submenú *CPLEX Parameters* del menú de *Options*, el usuario con experiencia puede incluso elegir entre una variedad de opciones la manera exacta de ejecutar el algoritmo para que se ajuste mejor al problema en cuestión.

Estas instrucciones para usar los paquetes de software se aclaran cuando se ve su aplicación en los ejemplos. Los archivos de Excel, LINGO/LINDO y MPL/CPLEX para este capítulo en el OR Courseware muestran cómo se aplican estas opciones con el ejemplo prototipo que se presenta en esta sección, y a otros ejemplos de PE subsecuentes.

La última parte del capítulo se centra en algoritmos de PE similares a los usados en estos paquetes. En la sección 11.6 se ilustra el algoritmo de PEB mediante la solución del ejemplo prototipo.

11.2 ALGUNAS APLICACIONES PEB

Igual que en el ejemplo de la California Manufacturing Co., con frecuencia los administradores enfrentan *decisiones de sí o no*. Por tanto, la *programación entera binaria* (PEB) se usa de manera considerable como ayuda para tomar estas decisiones.

Más adelante se introducirán varios tipos de decisiones sí o no. También se mencionan algunos ejemplos de aplicaciones reales donde se usó PEB para estudiar estas decisiones.

Cada una de estas aplicaciones se describe con detalle en un artículo de la revista *Interfaces*. Se hará hincapié en estos artículos que se incluyen en las referencias seleccionadas de aplicaciones que han sido premiadas que se citan al final del capítulo, puesto que se proporciona una liga a dichos artículos en el sitio en internet de este libro. En el caso de los demás artículos, se mencionará el número específico de la revista *Interfaces* en el que aparece el artículo.

Análisis de inversión

En ocasiones se usa programación lineal para tomar decisiones de presupuesto de capital acerca de cuánto invertir en diferentes proyectos. Sin embargo, como lo hace evidente el ejemplo de la California Manufacturing, algunas decisiones de presupuestos no se refieren a *cuánto* invertir sino

al hecho de *si debe* invertirse una cantidad fija. En especial, las cuatro decisiones del ejemplo eran si invertir una cantidad fija que es requerida para construir cierto tipo de instalación (fábrica o almacén) en determinado lugar (Los Ángeles o San Francisco).

La administración suele encontrarse en situaciones de decisión sobre si hacer una inversión fija (en las cuales la cantidad de capital se establece desde antes. ¿Debe adquirirse una subsidiaria en proceso de cierre de otra compañía? ¿Debe preferirse a cierto proveedor de materia prima? ¿Debe agregarse una nueva línea de producción para fabricar cierto subensamble en lugar de continuar con determinado proveedor?

En general, las decisiones de presupuestos de capital sobre inversiones fijas son decisiones sí o no del tipo siguiente.

Para cada decisión sí o no:
¿Debe realizarse cierta inversión fija?

$$\text{Variable de decisión} = \begin{cases} 1 & \text{si sí} \\ 0 & \text{si no.} \end{cases}$$

El número de julio-agosto de 1990 de *Interfaces* describe cómo *Turkish Petroleum Refineries Corporation* utiliza el enfoque PEB para analizar inversiones de capital del orden de decenas de millones de dólares para ampliar la capacidad de una refinería y ahorrar energía.

Un ejemplo distinto que todavía pertenece a esta categoría se describe en la referencia seleccionada A7. El enorme estudio de IO que se realizó para la *South African National Defense Force* para actualizar su capacidad con un presupuesto más pequeño. Las "inversiones" bajo consideración en este caso fueron los costos de adquisición y gastos de operación requeridos para proporcionar los tipos específicos de capacidad militar. Se formuló un modelo mixto de PEB con el objetivo de elegir las capacidades específicas para maximizar la eficacia global de la fuerza militar y satisfacer la restricción de presupuesto. El modelo tiene más de 16 mil variables —que incluyen 256 variables binarias— y más de 5 mil restricciones funcionales. La optimización acerca del tamaño y forma de la fuerza militar proporcionó ahorros de más de 1 100 millones de dólares anuales lo mismo que beneficios no monetarios. El efecto de este estudio hizo que ganara en 1996 el Franz Edelman Award por logros en Ciencias de la administración.

En una aplicación militar que de alguna forma es similar, la Comandancia Espacial de la Fuerza Aérea de Estados Unidos gasta muchos miles de millones de dólares cada año para adquirir y desarrollar vehículos y sistemas espaciales. El número de julio-agosto de 2003 de *Interfaces* describe la forma como la Comandancia Espacial emplea programación entera para optimizar estas inversiones a largo plazo en un horizonte de planeación de 24 años.

En la referencia seleccionada A3 se presenta otra aplicación premiada de un modelo de PEB mixto al análisis de inversiones. Este modelo particular fue utilizado por la compañía de inversiones *Grantham, Mayo, Van Otterloo and Company* para construir numerosos portafolios manejados de manera cuantitativa que representan más de 8 mil millones de dólares en títulos. En cada caso, se ha construido un portafolio parecido —en términos de sector y exposición de seguridad— a un portafolio objetivo pero con un número mucho más pequeño y manejable de acciones diferentes. Se utiliza una variable binaria para representar cada decisión del tipo sí o no para determinar si una acción particular se debe incluir en el portafolio; después, una variable continua representa la cantidad que se debe invertir en esa acción. Dado un portafolio que necesita ser rebalanceado, es deseable reducir los costos de transacción a través de minimización del número de operaciones necesarias para obtener el portafolio final, por lo que también es necesario incluir variables binarias para representar las decisiones del tipo sí o no para determinar si es necesario llevar a cabo las transacciones para cambiar las cantidades de las acciones que se posee. La inclusión de esta consideración en el modelo ha reducido el costo anual de transacciones de portafolios en al menos 4 millones de dólares.

Elección del sitio

En la economía globalizada, muchas corporaciones instalan nuevas plantas en diversas regiones del planeta para aprovechar la mano de obra más barata y otras ventajas. Antes de elegir la ubicación de esas instalaciones, deben analizarse y compararse muchos sitios potenciales. (El ejemplo de la California Manufacturing Co., tenía sólo dos sitios potenciales para cada una de dos tipos de instalaciones.) Cada ubicación potencial involucra una decisión sí o no del siguiente tipo.

Para cada decisión sí o no:
¿Debe elegirse cierto lugar para la ubicación de cierta instalación nueva?

$$\text{Variable de decisión} = \begin{cases} 1 & \text{si sí} \\ 0 & \text{si no.} \end{cases}$$

En muchos casos, el objetivo es seleccionar los sitios para minimizar el costo total de las nuevas instalaciones que proporcionarán la producción que se requiere.

Como se describe en la referencia seleccionada A11, AT&T utilizó un modelo de PEB para ayudar a docenas de clientes a seleccionar sitios para sus centros de telemercadeo. El modelo minimiza los costos de mano de obra, comunicaciones y bienes raíces a la vez que proporciona el nivel deseado de cobertura de los centros. En sólo un año (1988), este enfoque permitió a 46 clientes de AT&T tomar sus decisiones sí o no sobre las ubicaciones con presteza y confianza, al mismo tiempo que comprometieron $375 millones anuales en servicios de red y $31 millones en venta de equipo de AT&T.

A continuación describiremos un tipo importante de problemas en los que la selección del sitio tiene un papel esencial.

Diseño de una red de producción y distribución

Los fabricantes de hoy se enfrentan a una gran presión competitiva para lograr que sus productos lleguen al mercado con mayor rapidez, así como para reducir sus costos de producción y distribución. Por tanto, cualquier corporación que distribuye sus productos en un área geográfica grande —o incluso en todo el mundo— debe prestar una atención continua al diseño de sus productos y a la red de distribución.

Este diseño incluye investigar los siguientes tipos de decisiones sí o no.

¿Debe cierta planta permanecer abierta?
¿Debe seleccionarse cierto sitio para una nueva planta?
¿Debe cierto centro de distribución permanecer abierto?
¿Debe cierto sitio elegirse para instalar un nuevo centro de distribución?

Si cada área de mercado debe recibir servicio de un solo centro, entonces también es necesario tomar otro tipo de decisiones sí o no sobre cada combinación de área de mercado y centro de distribución.

¿Debe cierto centro de distribución asignarse a cierta área de mercado?

Por cada decisión sí o no de cualquiera de estos tipos:

$$\text{Variable de decisión} = \begin{cases} 1 & \text{si sí} \\ 0 & \text{si no.} \end{cases}$$

Ault Foods Limited (número de julio-agosto de 1994 de *Interfaces*) usó este enfoque para diseñar su centro de producción y distribución. La administración consideró 10 lugares para plantas, 13 lugares para centros de distribución y 48 áreas de mercado. Esta aplicación de PEB fue responsable de que la empresa pudiese ahorrar 200 mil dólares anuales.

Digital Equipment Corporation —este caso se describe en el número de enero-febrero de 1995 de *Interfaces*— es otro ejemplo de una aplicación de este tipo. En ese tiempo, esta gran corporación multinacional atendía 250 mil instalaciones de sus clientes, y la mitad de sus 14 mil millones de ingresos anuales provenían de 81 países fuera de Estados Unidos. Por tanto, esta aplicación consideraba la reestructuración de toda la *cadena de abastecimiento global* de la empresa, que consistía en proveedores, plantas, centros de distribución, sitios potenciales y áreas de mercado en todo el mundo. La reestructuración provocó reducciones de costos anuales por 500 millones en producción y 300 millones en logística, además de una reducción de más de 400 millones en activos de capital.

Despacho de envíos

Una vez que la red de producción y distribución haya sido diseñada y puesta en operación, deben tomarse decisiones operativas diarias acerca de cómo realizar los envíos. Algunas de estas decisiones también son de sí o no.

Por ejemplo, suponga que se usan camiones para transportar los envíos y que cada camión suele hacer entregas a varios clientes durante cada viaje. En consecuencia, es necesario elegir una ruta (secuencia de clientes) para cada camión, de manera que cada candidato para la ruta conduce a la siguiente decisión de sí o no.

¿Debe cierta ruta seleccionarse para uno de los camiones?

$$\text{Variable de decisión} = \begin{cases} 1 & \text{si sí} \\ 0 & \text{si no.} \end{cases}$$

El objetivo es seleccionar las rutas que minimizan el costo total de realizar los envíos.

También se pueden tomar en cuenta algunas complicaciones. Por ejemplo, si se dispone de camiones de varios tamaños, cada candidato para la elección debe incluir cierta ruta y cierto tamaño de camión. De igual manera, si el tiempo es un factor, se puede especificar el momento de salida como parte de la decisión de sí o no. Con ambos factores, cada decisión sí o no tendría la forma que se muestra a continuación.

Deben seleccionarse los siguientes elementos de manera simultánea para una entrega:

1. Cierta ruta,
2. Cierto tamaño de camión y
3. Cierto momento de salida

$$\text{Variable de decisión} = \begin{cases} 1 & \text{si sí} \\ 0 & \text{si no.} \end{cases}$$

Por ejemplo, *Sears, Roebuck and Company* —consulte el número de enero-febrero de 1999 de *Interfaces*— logró ahorros anuales por más de 42 millones de dólares mediante el empleo de un sistema de vehículos-rutas-y-programación basado en PEB y otro de información geográfica para realizar su servicio de entregas a domicilio de manera más eficiente.

Programación de actividades interrelacionadas

En la vida diaria, todas las personas programan actividades interrelacionadas, aun si sólo implican organizar el inicio de las tareas de la escuela. Los administradores también deben programar varios tipos de actividades interrelacionadas. ¿Cuándo se debe iniciar la producción de las nuevas órdenes? ¿Cuándo deben comercializarse los nuevos productos? ¿Cuándo deben realizarse las inversiones de capital para ampliar la capacidad de producción?

Para cualquier actividad como éstas, la decisión acerca de cuándo comenzar se puede expresar en términos de una serie de decisiones sí o no, una de ellas para cada momento posible de inicio, como se muestra en seguida.

¿Debe cierta actividad comenzar en un momento dado?

$$\text{Variable de decisión} = \begin{cases} 1 & \text{si sí} \\ 0 & \text{si no.} \end{cases}$$

Como una actividad en particular puede comenzar sólo en un momento, la elección de varios momentos proporciona un grupo de *alternativas mutuamente excluyentes*, de manera que la variable de decisión para uno solo de ellos puede tener un valor de 1.

Por ejemplo, considere la siguiente aplicación que ocurrió en *China* —número de enero-febrero de 1995 de *Interfaces*—. El país se enfrentaba a inversiones de al menos 240 mil millones de dólares en un horizonte de 15 años para cumplir con las necesidades de energía de su economía creciente. Para cubrir los faltantes de carbón y electricidad era necesario desarrollar una infraes-

Recuadro de aplicación

Las líneas aéreas comerciales deben resolver dos problemas complejos de programación de vuelos con el fin de garantizar que todos los vuelos que se programen tengan una tripulación disponible. Uno de ellos, llamado *problema de la planeación de viajes ocupados*, involucra la elaboración de secuencias de vuelo con periodos de descanso entre ellas que comprometerán los viajes ocupados por muchos días para las tripulaciones individuales. El segundo, que se llama *problema de las listas de turnos*, involucra la asignación de dichos viajes ocupados a cada uno de los miembros de la tripulación. La gerencia trata de encontrar soluciones de costo mínimo y máxima productividad a dichos problemas y así como satisfacer los acuerdos laborales y considerar las preferencias de los miembros de la tripulación.

Una gran cantidad de líneas aéreas alrededor del mundo han logrado ahorros significativos en años recientes mediante el uso de *modelos PEB* para obtener soluciones óptimas a dichos problemas. Una de estas aerolíneas es **Air New Zealand**, que es la línea aérea nacional e internacional más grande con base en este país. Emplea a más de 2 000 miembros de la tripulación y opera vuelos a Australia, Asia, Norteamérica y Europa, así como entre los centros más importantes de Nueva Zelanda.

Por lo general, los modelos PEB usados por la Air New Zealand cuentan con *cientos de restricciones funcionales y muchos miles de variables binarias*, en los que se utilizan técnicas avanzadas para resolver dichos modelos. Una estimación conservadora del ahorro que resulta del uso de estos modelos es de alrededor de **6.7 millones de dólares** *al año*, lo cual representa 11% de la ganancia operativa de la compañía en un año reciente. Asimismo, se presentaron muchos beneficios intangibles en los que se incluye una rápida implantación, la eficiente reprogramación de cambios de última hora y una mejora del servicio a los pasajeros.

Fuente: E. R. Butchers, P. R. Day, A. P. Goldie, S. Miller, J. A. Meyer, D. M. Ryan, A. C. Scott y C. A. Wallace: "Optimized Crew Scheduling at Air New Zealand", en *Interfaces*, **31**(1): 30-56, enero-febrero de 2001. (En el sitio en internet de este libro, www.mhhe.com/hillier, se proporciona una liga hacia este artículo.)

tructura para transportar carbón y transmitir energía, la construcción de nuevas presas y plantas para generar energía térmica, hidráulica y nuclear. Por tanto, la Comisión de Planeación del Gobierno de China y el Banco Mundial colaboraron para elaborar un gran modelo de PEB que debía guiar las decisiones acerca de qué proyectos aprobar y cuándo iniciarlos en un periodo de 15 años, para minimizar el costo descontado total. Se estima que esta aplicación de IO ha ahorrado a China alrededor de 6 400 millones de dólares en ese periodo.

Aplicaciones a líneas aéreas

La industria de líneas aéreas es un usuario constante de la IO en todas sus operaciones. Cientos de profesionales de la Investigación de Operaciones trabajan en esta área. Por lo general, las compañías aéreas más importantes tienen un departamento propio que trabaja en aplicaciones de IO. Además, hay algunas firmas de consultoría prominentes que se enfocan sólo en los problemas de compañías relacionadas con el transporte, en especial las líneas aéreas. Mencionaremos dos de las aplicaciones que se basan en PEB de manera directa.

Uno es el *problema de asignación de flota*. Dados varios modelos de aviones disponibles, el problema es asignar un tipo específico a cada tipo de trayecto de manera que se maximice la ganancia total si se cumple con lo programado. El trueque básico estriba en que si la línea aérea usa un avión demasiado pequeño en un trayecto dado, dejará clientes potenciales en tierra, mientras que si asigna uno demasiado grande, tendrá un gasto mucho mayor y volará con asientos vacíos.

Para cada combinación de tipo de avión y trayecto, se tiene la siguiente decisión sí o no.

¿Debe asignarse cierto tipo de avión a un trayecto en particular?

$$\text{Variable de decisión} = \begin{cases} 1 & \text{si sí} \\ 0 & \text{si no.} \end{cases}$$

Delta Air Lines realiza más de 2 500 vuelos nacionales todos los días, y emplea alrededor de 450 aviones de varios tipos. Como se describió en la referencia seleccionada A12, la empresa utiliza un modelo de programación lineal enorme —cerca de 40 mil restricciones funcionales, 20 mil variables binarias y 40 mil variables enteras— para resolver su problema de asignación de la flota cada vez que se requiere un cambio. Esta aplicación ahorra a Delta cerca de $100 millones anuales.

Una aplicación similar es el *problema de programación de tripulaciones*. En éste, en lugar de asignar tipos de aviones a rutas de vuelo, se asignan secuencias de rutas de vuelo a las tripulaciones

de pilotos y asistentes de vuelo. De esta forma, por cada secuencia factible de vuelos que sale de una base y regresa a la misma base, debe tomarse la siguiente decisión sí o no.

¿Debe cierta secuencia de rutas de vuelo asignarse a una tripulación?

$$\text{Variable de decisión} = \begin{cases} 1 & \text{si sí} \\ 0 & \text{si no.} \end{cases}$$

El objetivo es minimizar el costo total de proporcionar tripulaciones que cubran cada ruta según la programación.

American Airlines —este caso aparece en los números de julio-agosto de 1989 y enero-febrero de 1991 de *Interfaces*— logra ahorros anuales de más de $20 millones con el uso de PEB para resolver sus problemas de programación mensual de tripulaciones. Este enfoque también es muy usado por compañías aéreas que tienen sus oficinas centrales fuera de Estados Unidos. Por ejemplo, *Air New Zealand* —número de enero-febrero de 2001 de *Interfaces*— ahorra aproximadamente 6.7 millones de dólares al año debido al uso de PEB para optimizar la programación de tripulaciones, como se describió en el recuadro de aplicación de la página anterior.

Al final de la sección 11.4 se presentará un ejemplo de formulación completa de este tipo.

A las compañías aéreas se les presenta un problema similar cuando los programas de sus tripulaciones deben ser revisados de manera ocasional y con rapidez ante el acaecimiento de retrasos o cancelaciones de vuelos debidos a inclemencias del clima, problemas mecánicos en las naves o no disponibilidad de las tripulaciones. Como se describió en la Viñeta de aplicación de la sección 2.2 (así como en la referencia seleccionada A14), *Continental Airlines* logró ahorros de 40 millones de dólares en el primer año de uso de un sistema de apoyo a las decisiones basado en PEB para optimizar la *reasignación* de tripulaciones cuando ocurren tales emergencias. (En el año 2002, Continental Airlines ganó el primer lugar de los Franz Edelman Awards por los logros en ciencias de la administración por esta aplicación inovadora.)

Muchos de los problemas que enfrentan las compañías aéreas también surgen en otros segmentos de la industria del transporte. Por tanto, algunas de las aplicaciones de IO a los aviones se han extendido a estas otras áreas, incluso a los viajes en tren. Por ejemplo, tres de los ganadores al primer lugar de los Franz Edelman Award por logros en Ciencias de la administración en años recientes han sido por aplicaciones a la empresa nacional de ferrocarriles de Francia, la Canadian Pacific Railway y la Netherlands Railways que lograron beneficios financieros sorprendentes. (Vea los números de enero-febrero de 1998 y 2004 de *Interfaces*.) La referencia seleccionada A1 también describe la forma en que Netherlands Railways (NS Reizigers) ahorra en la actualidad 4.8 millones de dólares al año mediante el uso de BIP para optimizar la programación de su tripulación.

11.3 USOS INNOVADORES DE VARIABLES BINARIAS EN LA FORMULACIÓN DE MODELOS

Se han presentado varios ejemplos en los que las *decisiones básicas* del problema son del *tipo sí o no*, por lo que se introdujeron *variables binarias* para representarlas. Ahora se verán algunas otras formas en las que las variables binarias pueden ser útiles. En particular, se verá que estas variables a veces permiten tomar un problema cuya formulación no se puede manejar y *reformularlo* como un problema de PE pura o mixta.

Este tipo de situación surge cuando la formulación original del problema se ajusta a un problema de programación entera o a uno de programación lineal, *excepto* por ciertas diferencias menores que incluyen relaciones combinatorias en el modelo. Cuando se expresan estas relaciones en términos de preguntas cuya respuesta debe ser sí o no, se pueden introducir al modelo **variables binarias auxiliares** para representar estas decisiones sí o no. (En lugar de ser una variable de decisión del problema original en consideración, una variable binaria *auxiliar* es una variable binaria que se introduce en el modelo del problema con la finalidad de ayudar a formular el modelo como un modelo BIP puro o mezclado.) Al introducir estas variables, el problema se reduce a uno de PEM (o PE *pura* si todas las variables originales también están restringidas a valores enteros).

En seguida se presentan algunos casos que se pueden formular con este enfoque, donde las x_j denotan las variables *originales* del problema —pueden ser variables continuas o enteras—, y las y_i denotan las variables *auxiliares* que se introdujeron para realizar la nueva formulación.

Restricciones de tipo una u otra

Considere el importante caso en el que se debe elegir entre dos restricciones, de manera que *sólo una* (cualquiera de las dos) se tiene que cumplir (mientras que la otra puede cumplirse pero no es imprescindible que lo haga). Por ejemplo, puede existir la opción de usar uno de dos tipos de recursos para cierto propósito, de manera que sólo es necesario que una de las restricciones de disponibilidad de estos recursos se cumpla en forma matemática. Para ilustrar este tipo de situaciones, suponga que uno de los requerimientos en el problema completo es que

$$\begin{array}{ll} \text{se cumple} & 3x_1 + 2x_2 \leq 18 \\ \text{o} & x_1 + 4x_2 \leq 16, \end{array}$$

es decir, al menos una de estas dos desigualdades debe cumplirse pero no necesariamente las dos. Este requerimiento debe reformularse para que el problema se ajuste al formato de programación lineal, en el que se deben cumplir *todas* las restricciones especificadas. Si M es un número positivo muy grande, la expresión anterior se puede reescribir como

$$\begin{array}{ll} \text{se cumple} & \begin{array}{l} 3x_1 + 2x_2 \leq 18 \\ x_1 + 4x_2 \leq 16 + M \end{array} \\ \text{o} & \begin{array}{l} 3x_1 + 2x_2 \leq 18 + M \\ x_1 + 4x_2 \leq 16. \end{array} \end{array}$$

La clave es que al agregar M al lado derecho de una restricción de este tipo se obtiene el efecto de eliminarla, puesto que cualquier solución que satisfaga las otras restricciones del problema cumplirá ésta de manera automática. (Esta formulación supone que el conjunto de soluciones factibles del problema completo es un conjunto acotado y que M es lo suficientemente grande como para no eliminar ninguna de las soluciones factibles.) Esta formulación es equivalente al conjunto de restricciones

$$\begin{array}{l} 3x_1 + 2x_2 \leq 18 + My \\ x_1 + 4x_2 \leq 16 + M(1 - y). \end{array}$$

Debido a que la *variable auxiliar y* debe ser 0 o 1, esta formulación garantiza que una de las restricciones originales se cumple mientras que la otra es, de hecho, eliminada. Este nuevo conjunto de restricciones se podría añadir a las otras restricciones del modelo completo para obtener un problema de PE pura o mixta, lo cual depende de las x_j sean variables enteras o continuas.

Este enfoque se relaciona en forma directa con el análisis anterior sobre cómo expresar las relaciones combinatorias en términos de preguntas que deben responderse con sí o no. La relación combinatoria a la que se hace referencia concierne a la combinación de *otras* restricciones del modelo con la *primera* de las dos restricciones *alternas* y después con la *segunda*. ¿Cuál de estas dos combinaciones de restricciones es *mejor*, es decir, que se puede lograr en cuanto al valor de la función objetivo? Si esta pregunta se hace en términos de sí o no, en realidad es necesario formular dos preguntas complementarias:

1. ¿Debe elegirse $x_1 + 4x_2 \leq 16$ como la restricción que debe cumplirse?
2. ¿Debe elegirse $3x_1 + 2x_2 \leq 18$ como la restricción que debe cumplirse?

Como sólo una de estas dos preguntas tendrá una respuesta afirmativa, los respectivos términos binarios y y $1 - y$ representan las decisiones sí o no. En consecuencia, $y = 1$ si la respuesta a la primera pregunta es sí (y no a la segunda), mientras que $1 - y = 1$ (esto es $y = 0$) si la respuesta a la segunda pregunta es sí (y no a la primera). Como de manera automática $y + 1 - y = 1$ (un sí), no hay necesidad de agregar otra restricción para forzar a estas dos decisiones a que sean mutuamente excluyentes. (Si en lugar de este procedimiento se usan las variables binarias separadas y_1 y y_2 para representar estas decisiones sí o no, se necesitará una restricción adicional $y_1 + y_2 = 1$ para hacerlas mutuamente excluyentes.)

A continuación se ofrece una presentación formal de este enfoque para un caso más general.

Deben cumplirse *K* de *N* restricciones

Considere la situación en la que el modelo completo incluye un conjunto de N restricciones posibles entre las que sólo K de ellas se *deben* cumplir. (Suponga que $K < N$.) Parte del proceso de optimización es elegir la *combinación* de K restricciones que permita que la función objetivo adquiera el mejor valor posible. En realidad, las $N - K$ restricciones que *no* se eligen son eliminadas del problema, aun cuando por coincidencia las soluciones factibles puedan satisfacer algunas de ellas.

Este caso es una generalización directa del anterior que tenía $K = 1$ y $N = 2$. Denote las N restricciones posibles por

$$
\begin{aligned}
f_1(x_1, x_2, \ldots, x_n) &\leq d_1 \\
f_2(x_1, x_2, \ldots, x_n) &\leq d_2 \\
&\vdots \\
f_N(x_1, x_2, \ldots, x_n) &\leq d_N.
\end{aligned}
$$

Después, al aplicar la misma lógica que en el caso anterior, se encuentra que una formulación equivalente del requerimiento de que K de estas restricciones se *deben* cumplir es

$$
\begin{aligned}
f_1(x_1, x_2, \ldots, x_n) &\leq d_1 + My_1 \\
f_2(x_1, x_2, \ldots, x_n) &\leq d_2 + My_2 \\
&\vdots \\
f_N(x_1, x_2, \ldots, x_n) &\leq d_N + My_N \\
\sum_{i=1}^{N} y_i &= N - K,
\end{aligned}
$$

y

$$y_i \text{ es binaria}, \qquad \text{para } i = 1, 2, \ldots, N,$$

donde M es un número positivo muy grande. Para cada variable binaria y_i $(i = 1, 2, \ldots, N)$, observe que $y_i = 0$ hace que $My_i = 0$, lo que reduce la nueva restricción a la restricción original i. Por otro lado, $y_i = 1$ hace que $(d_i + My_i)$ sea tan grande que —si otra vez se supone una región factible acotada— cualquier solución que satisfaga las otras nuevas restricciones satisface de manera automática la nueva restricción i, lo cual tiene el efecto de eliminar la restricción original i. Por tanto, en razón de que las restricciones sobre las variables y_i garantizan que K de estas variables serán iguales a 0 y las restantes serán iguales a 1, K de las restricciones originales quedarán sin cambio y $(N - K)$ de ellas serán eliminadas. La elección de *cuáles* K de estas restricciones deben conservarse se hace mediante la aplicación del algoritmo apropiado al problema completo para encontrar una solución óptima para *todas* las variables de manera simultánea.

Funciones con *N* valores posibles

Considere la situación en la que se requiere que una función dada tome cualquiera de N valores dados. Denote este requisito por

$$f(x_1, x_2, \ldots, x_n) = d_1 \qquad \text{o} \qquad d_2, \ldots, \qquad \text{o} \qquad d_N.$$

Un caso especial es aquel en el que esta función es

$$f(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} a_j x_j,$$

como en el lado izquierdo de una restricción de programación lineal. Otro caso especial es en el que $f(x_1, x_2, \ldots, x_n) = x_j$ para un valor dado de j, de manera que el requerimiento en realidad es que x_j debe tomar cualquiera de los N valores dados.

La formulación equivalente de PE de este requerimiento es la siguiente:

$$f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{N} d_i y_i$$
$$\sum_{i=1}^{N} y_i = 1$$

y

y_i es binaria, para $i = 1, 2, \ldots, N$,

con lo que este nuevo conjunto de restricciones sustituye el requerimiento que se hizo al establecer el problema completo. Este conjunto proporciona una formulación *equivalente* puesto que exactamente una de las y_i debe ser igual a 1 y las otras deben ser iguales a 0, así que se escoge una d_i como el valor de la función. En este caso existen N preguntas con respuesta sí o no, que son: ¿debe ser d_i el valor escogido ($i = 1, 2, \ldots, N$)? Como la y_i respectiva representa estas *decisiones sí o no*, la segunda restricción las hace *alternativas mutuamente excluyentes*.

Para ilustrar cómo puede surgir un caso de este tipo, reconsidere el problema de la Wyndor Glass Co., que se presentó en la sección 3.1. En la actualidad no se emplean 18 horas de producción a la semana en la planta 3 y están disponibles para los dos nuevos productos *o* para otros que muy pronto estarán listos para producción. Con el fin de dejar cualquier capacidad restante en forma de bloques utilizables para estos productos, la administración desea imponer la restricción de que el tiempo de producción que se empleará en los dos nuevos productos actuales sea 6 *o* 12 *o* 18 horas semanales. Entonces, la tercera restricción del modelo original ($3x_1 + 2x_2 \leq 18$) se convierte en

$$3x_1 + 2x_2 = 6 \quad \text{o} \quad 12 \quad \text{o} \quad 18.$$

En la notación anterior, $N = 3$ con $d_1 = 6$, $d_2 = 12$ y $d_3 = 18$. En consecuencia, el nuevo requerimiento de la administración se debe formular de la siguiente manera:

$$3x_1 + 2x_2 = 6y_1 + 12y_2 + 18y_3$$
$$y_1 + y_2 + y_3 = 1$$

y

y_1, y_2, y_3 son binarias.

El modelo completo de esta nueva versión del problema consiste, entonces, en el modelo original (vea la sección 3.1), más este nuevo conjunto de restricciones que reemplaza a la tercera restricción original. Esta sustitución conduce a una formulación de PEM muy clara.

Problema de costo fijo

Es bastante común incurrir en un cargo de preparación o costo fijo cuando se emprende una actividad. Por ejemplo, un cargo así ocurre cuando se inicia una corrida de producción de un lote pequeño de un producto específico y deben prepararse las instalaciones fabriles que se requieren. En tales casos, el costo total de la actividad es la suma de un costo variable relacionado con el nivel de la actividad y el costo fijo en el que se incurre para iniciarla. Con frecuencia, el costo variable estará muy cerca de ser proporcional al nivel de la actividad. Si lo es, el *costo total* de la actividad (por ejemplo, j) se puede representar mediante una función de la forma

$$f_j(x_j) = \begin{cases} k_j + c_j x_j & \text{si } x_j > 0 \\ 0 & \text{si } x_j = 0, \end{cases}$$

donde x_j denota el nivel de la actividad j ($x_j \geq 0$), k_j denota el costo fijo y c_j el costo de cada unidad incremental. Si no fuera por el costo fijo k_j esta estructura de costo sugeriría la posibilidad de una formulación de *programación lineal* para determinar los niveles óptimos de las actividades competitivas. Por fortuna, aun si se incluyen las k_j, se puede usar programación entera.

Para formular el modelo completo, suponga que existen n actividades, cada una con la estructura de costo anterior —con $k_j \geq 0$ en todos los casos y $k_j > 0$ para alguna $j = 1, 2, \ldots, n$—, y que el problema es

$$\text{Minimizar} \qquad Z = f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n),$$

sujeta a

las restricciones de programación lineal dadas.

Para convertir este problema al formato de programación entera mixta, se comienza por proponer las preguntas que deban responderse con sí o no; por ejemplo, para cada $j = 1, 2, \ldots, n$, ¿debe emprenderse la actividad j ($x_j > 0$)? Después, cada una de estas decisiones sí o no se representa por una *variable binaria* y_j, de manera que

$$Z = \sum_{j=1}^{n} (c_j x_j + k_j y_j),$$

donde

$$y_j = \begin{cases} 1 & \text{si } x_j > 0 \\ 0 & \text{si } x_j = 0. \end{cases}$$

Por tanto, se puede pensar que las y_j son *decisiones contingentes* parecidas (pero no idénticas) a las que se consideran en la sección 11.1. Si M es un número positivo muy grande que excede el máximo valor factible de cualquiera de las x_j ($j = 1, 2, \ldots, n$), las restricciones

$$x_j \leq M y_j \qquad \text{para } j = 1, 2, \ldots, n$$

asegurarán que $y_j = 1$ y no cero, siempre que $x_j > 0$. La única dificultad que queda por salvar es que estas restricciones dejan a las y_j en libertad de tomar el valor 0 o 1 cuando $x_j = 0$. Por fortuna, esto también se resuelve en forma automática por la naturaleza de la función objetivo. El caso en el que $k_j = 0$ se puede ignorar puesto que y_j se puede eliminar de la formulación. Entonces, sólo se toma en cuenta el otro caso, es decir, cuando $k_j > 0$. Si $x_j = 0$, de manera que las restricciones permiten elegir entre $y_j = 0$ y $y_j = 1$, $y_j = 0$ debe conducir a un valor menor de Z que $y_j = 1$. De esta forma, como el objetivo es minimizar Z, un algoritmo que conduzca a una solución óptima siempre elegirá $y_j = 0$ si $x_j = 0$.

Para resumir, la formulación de PEM del problema de costo fijo es

$$\text{Minimizar} \qquad Z = \sum_{j=1}^{n} (c_j x_j + k_j y_j),$$

sujeta

las restricciones originales, más
$x_j - My_j \leq 0$

y

y_j es binaria, para $j = 1, 2, \ldots, n$.

Si las x_j también estuvieran restringidas a valores enteros, éste sería un problema de PE *pura*.

Para ilustrar este enfoque, vea de nuevo el problema de la contaminación ambiental que enfrenta la Nori & Leets Co., que se presentó en la sección 3.4. El primer método de reducción que se consideró —aumentar la altura de las chimeneas— en realidad significaría un enorme *costo fijo* para poder realizar *cualquier* aumento, además de un costo variable que en esencia sería proporcional al aumento. Después de una conversión al costo anual equivalente que se utilizó en la formulación, este cargo fijo sería de 2 millones de dólares por cada uno de los altos hornos y los hornos de corazón abierto, mientras que los costos variables son los mismos que se identificaron en la tabla 3.14. Por tanto, en la notación anterior, $k_1 = 2$, $k_2 = 2$, $c_1 = 8$ y $c_2 = 10$, donde la función objetivo se expresa en unidades de *millones* de dólares. Como los otros métodos de reducción no

implican cargos fijos, $k_j = 0$ para $j = 3, 4, 5, 6$. En consecuencia, la nueva formulación de PEM de este problema es

$$\text{Minimizar} \qquad Z = 8x_1 + 10x_2 + 7x_3 + 6x_4 + 11x_5 + 9x_6 + 2y_1 + 2y_2,$$

sujeta a

las restricciones dadas en la sección 3.4, más

$$x_1 - My_1 \leq 0,$$
$$x_2 - My_2 \leq 0,$$

y

y_1, y_2 son binarias.

Representación binaria de variables enteras en general

Suponga que se tiene un problema de PE pura en el que la mayoría de las variables son *binarias*, pero la presencia de algunas variables enteras *generales* evita que se pueda usar el eficiente algoritmo de PEB ahora disponible. Una manera de salvar esta dificultad es usar la *representación binaria* de cada una de estas variables enteras generales. En particular, si las cotas de una variable entera x son

$$0 \leq x \leq u$$

y si N está definido como un entero tal que

$$2^N \leq u < 2^{N+1},$$

entonces la **representación binaria** de x es

$$x = \sum_{i=0}^{N} 2^i y_i,$$

donde las variables y_i son variables binarias (auxiliares). Si se sustituye esta representación binaria por cada variable entera general —con un conjunto diferente de variables binarias auxiliares de cada una—, el problema completo se reduce a un modelo de PEB.

Por ejemplo, suponga que un problema de PE tiene sólo dos variables enteras generales, x_1 y x_2 y muchas variables binarias. Suponga también que el problema tiene restricciones de no negatividad para ambas variables y que las restricciones funcionales incluyen

$$\begin{aligned} x_1 \qquad &\leq 5 \\ 2x_1 + 3x_2 &\leq 30. \end{aligned}$$

Estas restricciones implican que $u = 5$ para x_1 y $u = 10$ para x_2, de manera que la definición anterior de N da $N = 2$ para x_1 —puesto que $2^2 \leq 5 < 2^3$— y $N = 3$ para x_2 (porque $2^3 \leq 10 < 2^4$). Entonces, las representaciones binarias de estas variables son

$$\begin{aligned} x_1 &= y_0 + 2y_1 + 4y_2 \\ x_2 &= y_3 + 2y_4 + 4y_5 + 8y_6. \end{aligned}$$

Después de sustituir estas expresiones por las variables respectivas en todas las restricciones funcionales y en la función objetivo, las dos restricciones funcionales establecidas son:

$$\begin{aligned} y_0 + 2y_1 + 4y_2 \qquad\qquad\qquad\qquad\qquad\qquad &\leq 5 \\ 2y_0 + 4y_1 + 8y_2 + 3y_3 + 6y_4 + 12y_5 + 24y_6 &\leq 30. \end{aligned}$$

Observe que cada valor factible de x_1 corresponde a uno de los valores factibles del vector (y_0, y_1, y_2), lo cual también ocurre con x_2 y (y_3, y_4, y_5, y_6). Por ejemplo, $x_1 = 3$ corresponde a $(y_0, y_1, y_2) = (1, 1, 0)$ y $x_2 = 5$ corresponde a $(y_3, y_4, y_5, y_6) = (1, 0, 1, 0)$.

Si se trata de un problema de PE en el que *todas* las variables son enteras generales (acotadas), es posible usar esta misma técnica para reducirlo a un modelo de PEB. Sin embargo, esto casi nunca

es aconsejable debido a la explosión del número de variables involucradas. En general, aplicar un buen algoritmo de PE al modelo original será más eficiente que aplicar un buen algoritmo de PEB al modelo mucho más grande de PEB.[1]

En términos generales, con *todas* las posibilidades de formulación con variables binarias auxiliares que se presentan en esta sección, es necesario hacer notar una precaución que debe tenerse. Algunas veces, este enfoque requiere que se agregue un número relativamente grande de variables, lo que puede hacer que el modelo se vuelva *no factible en cuanto a los cálculos*. (En la sección 11.5 se proporcionan algunas perspectivas sobre los tamaños de los problemas de PE que se pueden resolver.)

11.4 ALGUNOS EJEMPLOS DE FORMULACIÓN

A continuación se presentará una serie de ejemplos que ilustran varias técnicas de formulación con variables binarias, entre ellos los que se estudiaron en las dos secciones anteriores. Por claridad, estos ejemplos se han mantenido muy pequeños. (En el sitio en internet de este libro, dentro de la sección de Worked Examples, se incluye un **ejemplo de formulación más grande** con docenas de variables binarias y restricciones). En las aplicaciones reales, estas formulaciones son sólo una pequeña parte de un modelo mucho más amplio.

EJEMPLO 1 Elecciones cuando las variables de decisión son continuas

La división de investigación y desarrollo de la GOOD PRODUCTS COMPANY ha desarrollado tres nuevos productos posibles. Sin embargo, para evitar una diversificación excesiva de la línea de productos de la compañía, la administración ha impuesto las siguientes limitaciones.

Restricción 1: De los tres nuevos productos posibles, deben escogerse, *como máximo*, sólo *dos* de ellos.

Se dispone de dos plantas que pueden fabricar los productos elegidos. Por razones administrativas, la administración impuso una segunda restricción a este respecto.

Restricción 2: Sólo una de las dos plantas debe asignarse para la producción de los nuevos productos.

En esencia, el costo unitario de producción de cada producto sería el mismo en las dos plantas. Sin embargo, por diferencias en las instalaciones, el número de horas de producción por unidad de cada producto puede diferir entre ellas. Estos datos se dan en la tabla 11.2, junto con otra información relevante, que incluye las estimaciones del departamento de mercadotecnia sobre el número máximo de unidades de cada producto que podrían venderse a la semana. El objetivo es seleccionar los productos, la planta y las tasas de producción de los bienes elegidos de manera que se maximice la ganancia total.

TABLA 11.2 Datos del ejemplo 1 (problema de la Good Products Co.)

	Tiempo de producción utilizado por cada unidad producida			Horas disponibles por semana
	Producto 1	Producto 2	Producto 3	
Planta 1	3 horas	4 horas	2 horas	30 horas
Planta 2	4 horas	6 horas	2 horas	40 horas
Ganancia unitaria	5	7	3	(miles de dólares)
Ventas potenciales	7	5	9	(unidades por semana)

[1] Se puede consultar una evidencia que soporta esta conclusión en J. H. Owen y S. Mehrotra, "On the Value of Binary Expansions for General Mixed Integer Linear Programs", en *Operations Research*, **50**: 810-819, 2002.

De alguna manera este problema se parece al *problema de mezcla de productos* estándar del ejemplo de la Wyndor Glass Co., que se describió en la sección 3.1. En realidad, si se eliminaran las dos restricciones anteriores y se requiriera que cada unidad de producto usara las horas de producción que se proporcionan en la tabla 11.2 en *ambas plantas*, se convertiría en un problema tipo. En particular, sean x_1, x_2 y x_3 las tasas de producción de los respectivos productos, el modelo sería

$$\text{Maximizar} \quad Z = 5x_1 + 7x_2 + 3x_3,$$

sujeta a

$$\begin{aligned} 3x_1 + 4x_2 + 2x_3 &\le 30 \\ 4x_1 + 6x_2 + 2x_3 &\le 40 \\ x_1 \qquad\qquad &\le 7 \\ x_2 \qquad &\le 5 \\ x_3 &\le 9 \end{aligned}$$

y

$$x_1 \ge 0, \qquad x_2 \ge 0, \qquad x_3 \ge 0.$$

No obstante, cuando se trata de problemas reales, la limitación 1 necesita que se agregue al modelo la siguiente restricción:

El número de variables de decisión estrictamente positivas (x_1, x_2, x_3) debe ser ≤ 2.

Esta restricción no se ajusta al formato de programación lineal o entera, por lo que la pregunta clave es cómo convertirla al formato que permita usar un algoritmo eficiente para resolver el modelo completo. Si las variables de decisión fueran binarias, entonces la restricción se expresaría como $x_1 + x_2 + x_3 \le 2$. Sin embargo, con variables de decisión *continuas*, se necesita un enfoque más complicado que incluya la introducción de variables binarias auxiliares.

La limitación 2 necesita que se sustituyan las dos primeras restricciones funcionales ($3x_1 + 4x_2 + 2x_3 \le 30$ y $4x_1 + 6x_2 + 2x_3 \le 40$) por la restricción

$$\begin{aligned} &\text{ya sea} \quad 3x_1 + 4x_2 + 2x_3 \le 30 \\ &\text{o bien} \quad 4x_1 + 6x_2 + 2x_3 \le 40 \end{aligned}$$

debe cumplirse, donde la elección de cuál restricción se cumple corresponde a la elección de qué planta se usará para producir los nuevos productos. En la sección anterior se analizó cómo se puede convertir una restricción de una o la otra al formato de programación lineal o entera, de nuevo con la ayuda de variables binarias auxiliares.

Formulación con variables binarias auxiliares. Para manejar la limitación 1, se introducen tres variables binarias auxiliares (y_1, y_2, y_3) con la interpretación

$$y_j = \begin{cases} 1 & \text{si } x_j > 0 \text{ se puede cumplir (se puede producir } j) \\ 0 & \text{si } x_j = 0 \text{ debe cumplirse (no se puede producir } j), \end{cases}$$

para $j = 1, 2, 3$. Para incluir esta interpretación en el modelo con la ayuda de M —un número positivo muy grande—, se agregan las restricciones

$$\begin{aligned} &x_1 \le My_1 \\ &x_2 \le My_2 \\ &x_3 \le My_3 \\ &y_1 + y_2 + y_3 \le 2 \\ &y_j \text{ es binaria,} \qquad \text{para } j = 1, 2, 3. \end{aligned}$$

La restricción de una o la otra y las restricciones de no negatividad proporcionan una región *acotada* para las variables de decisión, con lo que cada $x_j \le M$ en toda la región. Por tanto, para cada $x_j \le My_j$ la restricción $y_j = 1$ permite que x_j tome cualquier valor en la región factible, mientras que $y_j = 0$ fuerza a que $x_j = 0$. (De manera inversa, $x_j > 0$ hace que $y_j = 1$, mientras que $x_j = 0$ permite que y_j tome cualquiera de los dos valores.) En consecuencia, cuando la cuarta restricción hace que

a lo más dos de las y_j sean iguales a 1, significa que es posible elegir, cuando mucho, fabricar dos de los nuevos productos.

Para manejar la limitación 2 se introduce otra variable binaria auxiliar y_4 con la interpretación

$$y_4 = \begin{cases} 1 & \text{si } 4x_1 + 6x_2 + 2x_3 \leq 40 \text{ debe cumplirse (se elige la planta 2)} \\ 0 & \text{si } 3x_1 + 4x_2 + 2x_3 \leq 30 \text{ debe cumplirse (se elige la planta 1).} \end{cases}$$

Como se vio en la sección 11.3, esta interpretación se toma en cuenta al agregar las restricciones,

$$3x_1 + 4x_2 + 2x_3 \leq 30 + My_4$$
$$4x_1 + 6x_2 + 2x_3 \leq 40 + M(1 - y_4)$$
$$y_4 \text{ es binaria.}$$

En consecuencia, después de mover todas las variables al lado izquierdo de las restricciones, el modelo completo es

$$\text{Maximizar} \quad Z = 5x_1 + 7x_2 + 3x_3,$$

sujeta a

$$\begin{aligned} x_1 &\leq 7 \\ x_2 &\leq 5 \\ x_3 &\leq 9 \\ x_1 - My_1 &\leq 0 \\ x_2 - My_2 &\leq 0 \\ x_3 - My_3 &\leq 0 \\ y_1 + y_2 + y_3 &\leq 2 \\ 3x_1 + 4x_2 + 2x_3 - My_4 &\leq 30 \\ 4x_1 + 6x_2 + 2x_3 + My_4 &\leq 40 + M \end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0$$
$$y_j \text{ es binaria,} \quad \text{para } j = 1, 2, 3, 4.$$

Éste es ahora un modelo de PEM, con tres variables (las x_j) que no tienen que ser enteras y cuatro variables binarias, por lo cual es posible usar un algoritmo de PEM para resolverlo. Luego de ello —después de sustituir M por un valor numérico grande—,[2] la solución óptima es $y_1 = 1$, $y_2 = 0$, $y_3 = 1$, $y_4 = 1$, $x_1 = 5\frac{1}{2}$, $x_2 = 0$ y $x_3 = 9$; es decir, se eligen los productos 1 y 3 para producción, se elige la planta 2 y las tasas de producción deben ser $5\frac{1}{2}$ unidades por semana del producto 1 y 9 unidades por semana del producto 3. La ganancia total resultante es de \$54 500 semanales.

EJEMPLO 2 Violación de la proporcionalidad

La CORPORACIÓN SUPERSUDS ha comenzado a desarrollar sus planes de comercialización de sus nuevos productos para el próximo año. Considera la posibilidad de comprar un total de cinco comerciales de televisión en las redes nacionales para promocionar tres de estos productos. El problema que se debe resolver es cómo asignar los cinco comerciales a estos productos, con un máximo de tres —y un mínimo de cero— para cada uno de ellos.

En la tabla 11.3 se muestra el efecto estimado de asignar 0, 1, 2 o 3 comerciales a cada producto. Este efecto se mide en términos de la *ganancia* (en millones de dólares) que será resultado de las *ventas adicionales* generadas por los comerciales. El objetivo es asignar cinco comerciales a los productos de manera que se maximice la ganancia total.

[2] En la práctica, debe tenerse cuidado de dar un valor a M que sea lo suficientemente grande como para evitar que se eliminen soluciones factibles, pero tan pequeño como sea posible para evitar una región factible innecesariamente grande para el relajamiento de PL (descrito en la siguiente sección) y para evitar inestabilidad numérica. El examen cuidadoso de las restricciones en este ejemplo revela que el valor factible mínimo de M es $M = 9$.

Este pequeño problema se puede resolver con facilidad con programación dinámica (capítulo 10) o incluso por inspección. (La solución óptima es asignar dos comerciales al producto 1, ninguno al producto 2 y 3 al producto 3). Sin embargo, como ilustración, se desarrollarán dos formulaciones diferentes de PEB. Este tipo de formulación sería necesaria si este pequeño problema tuviera que incorporarse a un modelo de PE más grande que tal vez involucrara la asignación de recursos a las actividades de comercialización de los nuevos productos de toda la corporación.

Una formulación con variables binarias auxiliares. Una formulación natural surge si se denota por x_1, x_2, x_3 el número de comerciales de TV asignados a los respectivos productos. En consecuencia, la contribución de cada x_j a la función objetivo se presenta en la columna correspondiente de la tabla 11.3. Sin embargo, cada una de estas columnas viola el supuesto de proporcionalidad descrito en la sección 3.3, debido a lo cual no se puede establecer una función objetivo *lineal* en términos de estas variables de decisión enteras.

A continuación, observe lo que ocurre cuando se introducen *variables binarias auxiliares* y_{ij} en cada valor entero positivo de $x_i = j$ ($j = 1, 2, 3$), donde y_{ij} se interpreta como

$$y_{ij} = \begin{cases} 1 & \text{si } x_i = j \\ 0 & \text{de otra manera.} \end{cases}$$

(Por ejemplo, $y_{21} = 0$, $y_{22} = 0$ y $y_{23} = 1$, significa que $x_2 = 3$.) El modelo de PEB *lineal* que se obtiene es

$$\text{Maximizar} \quad Z = y_{11} + 3y_{12} + 3y_{13} + 2y_{22} + 3y_{23} - y_{31} + 2y_{32} + 4y_{33},$$

sujeta a

$$\begin{aligned} y_{11} + y_{12} + y_{13} &\le 1 \\ y_{21} + y_{22} + y_{23} &\le 1 \\ y_{31} + y_{32} + y_{33} &\le 1 \\ y_{11} + 2y_{12} + 3y_{13} + y_{21} + 2y_{22} + 3y_{23} + y_{31} + 2y_{32} + 3y_{33} &= 5 \end{aligned}$$

y

todas las y_{ij} es binaria.

Observe que las tres primeras restricciones funcionales aseguran que a cada x_i se le asignará exactamente uno de sus valores posibles. (Aquí, $y_{i1} + y_{i2} + y_{i3} = 0$ corresponde a $x_i = 0$, lo que no contribuye en nada a la función objetivo.) La última restricción funcional asegura que $x_1 + x_2 + x_3 = 5$. En consecuencia, de acuerdo con la tabla 11.3, la función objetivo *lineal* proporciona la ganancia total.

Al resolver este modelo de PEB se obtiene la siguiente solución óptima

$$\begin{aligned} &y_{11} = 0, \quad y_{12} = 1, \quad y_{13} = 0, \quad \text{es decir,} \quad x_1 = 2 \\ &y_{21} = 0, \quad y_{22} = 0, \quad y_{23} = 0, \quad \text{es decir,} \quad x_2 = 0 \\ &y_{31} = 0, \quad y_{32} = 0, \quad y_{33} = 1, \quad \text{es decir,} \quad x_3 = 3. \end{aligned}$$

TABLA 11.3 Datos del ejemplo 2 (problema de la Supersuds Corp.)

	Ganancia		
	Producto		
Número de comerciales	1	2	3
0	0	0	0
1	1	0	−1
2	3	2	2
3	3	3	4

Otra formulación con variables binarias auxiliares. Se vuelven a definir las variables binarias auxiliares anteriores y_{ij} de la siguiente manera:

$$y_{ij} = \begin{cases} 1 & \text{si } x_i \geq j \\ 0 & \text{de otra manera.} \end{cases}$$

Entonces, la diferencia estriba en que ahora $y_{ij} = 1$ si $x_i \geq j$ en lugar de que $x_i = j$. Entonces,

$$\begin{aligned} x_i = 0 \quad &\Rightarrow \quad y_{i1} = 0, \quad y_{i2} = 0, \quad y_{i3} = 0, \\ x_i = 1 \quad &\Rightarrow \quad y_{i1} = 1, \quad y_{i2} = 0, \quad y_{i3} = 0, \\ x_i = 2 \quad &\Rightarrow \quad y_{i1} = 1, \quad y_{i2} = 1, \quad y_{i3} = 0, \\ x_i = 3 \quad &\Rightarrow \quad y_{i1} = 1, \quad y_{i2} = 1, \quad y_{i3} = 1, \end{aligned}$$
de manera que $x_i = y_{i1} + y_{i2} + y_{i3}$

para $i = 1, 2, 3$. Se refuerzan las definiciones de que se permite que $y_{i2} = 1$ sea contingente respecto de $y_{i1} = 1$, y que $y_{i3} = 1$ lo es respecto de $y_{i2} = 1$, por medio las restricciones

$$y_{i2} \leq y_{i1} \quad \text{y} \quad y_{i3} \leq y_{i2}, \quad \text{para } i = 1, 2, 3.$$

La nueva definición de y_{ij} cambia también la función objetivo, como se ilustra en la figura 11.1, en la parte que corresponde al producto 1 de esta función. Como y_{11}, y_{12} y y_{13} proporcionan los incrementos sucesivos (si los hay) del valor de x_1 (a partir de un valor de 0), los coeficientes de y_{11}, y_{12} y y_{13} están dados por los *incrementos* respectivos en la columna del producto 1 de la tabla 11.3 ($1 - 0 = 1$, $3 - 1 = 2$, $3 - 3 = 0$). Estos *incrementos* son las *pendientes* en la figura 11.1, que conducen a $1y_{11} + 2y_{12} + 0y_{13}$ para la parte del producto 1 en la función objetivo. Observe que la aplicación de este enfoque a los tres productos conduce a una función objetivo *lineal*.

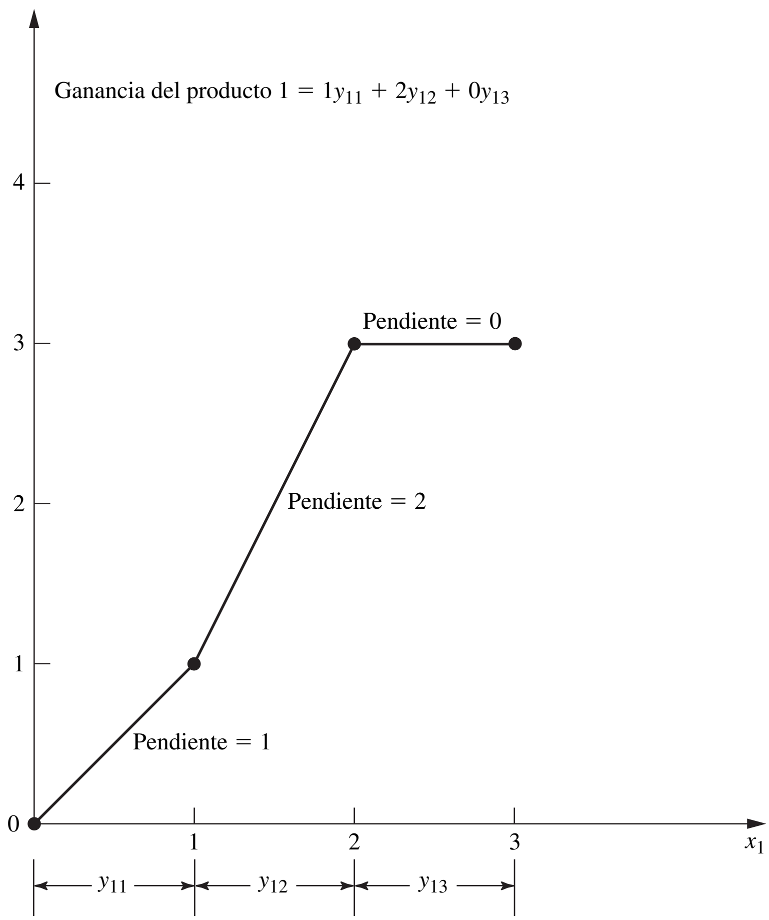


FIGURA 11.1
Utilidad obtenida por las ventas adicionales del producto 1 generada por x_1 anuncios de televisión, donde las pendientes dadas corresponden a los coeficientes de la función objetivo de la segunda formulación de PEB del ejemplo 2 (problema de la Supersuds Corp.).

Después de mover todas las variables al lado izquierdo de las restricciones el modelo completo de PEB que se obtiene es el siguiente:

Maximizar $\quad Z = y_{11} + 2y_{12} + 2y_{22} + y_{23} - y_{31} + 3y_{32} + 2y_{33}$,

sujeta a

$$\begin{aligned} y_{12} - y_{11} &\le 0 \\ y_{13} - y_{12} &\le 0 \\ y_{22} - y_{21} &\le 0 \\ y_{23} - y_{22} &\le 0 \\ y_{32} - y_{31} &\le 0 \\ y_{33} - y_{32} &\le 0 \\ y_{11} + y_{12} + y_{13} + y_{21} + y_{22} + y_{23} + y_{31} + y_{32} + y_{33} &= 5 \end{aligned}$$

y

todas las y_{ij} son binarias.

Al resolver este modelo de PEB se obtiene la siguiente solución óptima

$$\begin{array}{lllll} y_{11} = 1, & y_{12} = 1, & y_{13} = 0, & \text{es decir,} & x_1 = 2 \\ y_{21} = 0, & y_{22} = 0, & y_{23} = 0, & \text{es decir,} & x_2 = 0 \\ y_{31} = 1, & y_{32} = 1, & y_{33} = 1, & \text{es decir,} & x_3 = 3. \end{array}$$

Existen pocos argumentos para elegir entre este modelo de PEB y el anterior excepto el gusto personal. Tienen el mismo número de variables binarias, una consideración primordial cuando se determina el esfuerzo computacional en este tipo de problemas. Ambos tienen alguna *estructura especial* —restricciones para *alternativas mutuamente excluyentes* en el primer modelo y restricciones para las *decisiones contingentes* en el segundo— que puede ayudar a resolverlo con mayor rapidez. El segundo modelo tiene más restricciones funcionales que el primero.

EJEMPLO 3 Cobertura de todas las características

SOUTHWESTERN AIRWAYS necesita asignar sus tripulaciones para cubrir todos sus vuelos programados. Se estudiará el problema de asignar tres tripulaciones con base en San Francisco (SF) a los vuelos enumerados en la tabla 11.4. Las otras 12 columnas muestran 12 secuencias de vuelos factibles de una tripulación. (Los números en cada columna indican el orden de los vuelos.) Es necesario elegir tres de estas secuencias (una por tripulación) de tal manera que se cubran todos los vuelos. (Se permite tener más de una tripulación en un vuelo, en el cual los miembros de la tripulación adicional volarían como pasajeros, pero los contratos colectivos de trabajo requieren que se pague el tiempo de la tripulación adicional como si estuviera en horario de trabajo.) El costo de asignar una tripulación a una secuencia de vuelos específica se muestra (en miles de dólares) en el renglón inferior de la tabla. El objetivo es minimizar el costo total de asignar las tres tripulaciones de manera que cubran todos los vuelos.

Formulación con variables binarias. Con 12 secuencias factibles de vuelos, se deben enfrentar 12 decisiones sí o no:

¿Debe asignarse la secuencia j a la tripulación? $\quad (j = 1, 2, \ldots, 12)$

Por tanto, se usan 12 variables binarias para representar las respectivas decisiones:

$$x_j = \begin{cases} 1 & \text{si se asigna la secuencia } j \text{ a una tripulación} \\ 0 & \text{de otra manera.} \end{cases}$$

La parte más interesante de esta formulación es la naturaleza de cada restricción que asegura que se cubra el vuelo correspondiente. Por ejemplo, considere el último vuelo de la tabla 11.4 [Seattle a Los Ángeles (LA)]. Cinco secuencias —es decir, las secuencias 6, 9, 10, 11 y 12— in-

TABLA 11.4 Datos del ejemplo 3 (el problema de la aerolínea Southwestern)

	Secuencia factible de vuelos											
Vuelo	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1. San Francisco a Los Ángeles	1			1			1			1		
2. San Francisco a Denver		1			1			1			1	
3. San Francisco a Seattle			1			1			1			1
4. Los Ángeles a Chicago				2			2		3	2		3
5. Los Ángeles a San Francisco	2					3				5	5	
6. Chicago a Denver				3	3				4			
7. Chicago a Seattle							3	3		3	3	4
8. Denver a San Francisco		2		4	4				5			
9. Denver a Chicago					2			2			2	
10. Seattle a San Francisco			2				4	4				5
11. Seattle a Los Ángeles						2			2	4	4	2
Costo, en miles de dólares	2	3	4	6	7	5	7	8	9	9	8	9

cluyen este vuelo. Por tanto, debe elegirse al menos una de estas secuencias. La restricción que resulta es

$$x_6 + x_9 + x_{10} + x_{11} + x_{12} \geq 1.$$

Con restricciones similares para los otros 10 vuelos, el modelo completo de PEB es

$$\text{Minimizar} \quad Z = 2x_1 + 3x_2 + 4x_3 + 6x_4 + 7x_5 + 5x_6 + 7x_7 + 8x_8 + 9x_9 + 9x_{10} + 8x_{11} + 9x_{12},$$

sujeta a

$$\begin{aligned}
x_1 + x_4 + x_7 + x_{10} &\geq 1 && \text{(SF a LA)}\\
x_2 + x_5 + x_8 + x_{11} &\geq 1 && \text{(SF a Denver)}\\
x_3 + x_6 + x_9 + x_{12} &\geq 1 && \text{(SF a Seattle)}\\
x_4 + x_7 + x_9 + x_{10} + x_{12} &\geq 1 && \text{(LA a Chicago)}\\
x_1 + x_6 + x_{10} + x_{11} &\geq 1 && \text{(LA a SF)}\\
x_4 + x_5 + x_9 &\geq 1 && \text{(Chicago a Denver)}\\
x_7 + x_8 + x_{10} + x_{11} + x_{12} &\geq 1 && \text{(Chicago a Seattle)}\\
x_2 + x_4 + x_5 + x_9 &\geq 1 && \text{(Denver a SF)}\\
x_5 + x_8 + x_{11} &\geq 1 && \text{(Denver a Chicago)}\\
x_3 + x_7 + x_8 + x_{12} &\geq 1 && \text{(Seattle a SF)}\\
x_6 + x_9 + x_{10} + x_{11} + x_{12} &\geq 1 && \text{(Seattle a LA)}\\
\sum_{j=1}^{12} x_j &= 3 && \text{(asignar tres tripulaciones)}
\end{aligned}$$

y

$$x_j \text{ es binaria,} \qquad \text{para } j = 1, 2, \ldots, 12.$$

Una solución óptima para este modelo de PEB es

$$\begin{aligned}
x_3 &= 1 && \text{(asignar la secuencia 3 a una tripulación)}\\
x_4 &= 1 && \text{(asignar la secuencia 4 a una tripulación)}\\
x_{11} &= 1 && \text{(asignar la secuencia 11 a una tripulación)}
\end{aligned}$$

y todas las demás $x_j = 0$, con un costo total de \$18 000 (otra solución óptima es $x_1 = 1$, $x_5 = 1$, $x_{12} = 1$, y todas las demás $x_j = 0$).

Este ejemplo ilustra una clase más amplia de problemas llamada **problemas de cobertura de conjunto**.[3] Cualquier problema de cobertura se puede describir en términos generales como uno que involucra a cierto número de *actividades* potenciales (como las secuencias de vuelo) y *características* (como los vuelos). Cada actividad posee alguna pero no todas las características. El objetivo es determinar la combinación menos costosa de actividades que de manera colectiva posea (cubra) todas las características, por lo menos una vez. Así, si S_i es el conjunto de todas las actividades que poseen la característica i, al menos un elemento de S_i debe quedar incluido entre las actividades seleccionadas, por lo que la restricción,

$$\sum_{j \in S_i} x_j \geq 1,$$

se incluye en cada característica i.

Una clase de problemas relacionada, conocida como **problemas de partición de conjuntos**, cambia todas las restricciones de este tipo a

$$\sum_{j \in S_i} x_j = 1,$$

por lo que debe incluirse *exactamente* un elemento de cada conjunto S_i entre las actividades seleccionadas. En el ejemplo de la programación de tripulaciones, esto significa que cada vuelo debe estar incluido exactamente una vez entre las secuencias elegidas, lo que elimina la posibilidad de llevar tripulaciones adicionales (como pasajeros) en cualquier vuelo.

11.5 ALGUNAS PERSPECTIVAS ACERCA DE LA SOLUCIÓN DE PROBLEMAS DE PROGRAMACIÓN ENTERA

Puede parecer que la solución de problemas de PE es sencilla. Después de todo, los problemas de *programación lineal* se pueden resolver en forma eficiente, y la única diferencia es que la PE tiene muchas menos soluciones que considerar. En realidad, se garantiza que los problemas de PE *pura* con región factible acotada tengan sólo un número *finito* de soluciones factibles.

Desafortunadamente, existen dos falacias en esta línea de razonamiento. Una es que tener un número finito de soluciones factibles asegura que el problema se puede resolver. Los números finitos pueden ser astronómicamente grandes. Por ejemplo, considere el caso sencillo de problemas de PEB. Si se tienen n variables, existen 2^n soluciones que considerar (algunas de las cuales se pueden descartar por violar las restricciones funcionales). En consecuencia, cada vez que n aumenta en 1, el número de soluciones se *duplica*. Este patrón se llama **crecimiento exponencial** de la dificultad del problema. Con $n = 10$, existen más de mil soluciones (1 024); con $n = 20$, son más de un millón; con $n = 30$ resultan más de mil millones, y así sucesivamente; en razón de ello, aun las computadoras más eficientes son incapaces de realizar una enumeración exhaustiva —que verifique la factibilidad de cada solución y, de ser factible, que calcule el valor de la función objetivo— en problemas de PEB con unas cuantas docenas de variables, sin mencionar los problemas de PE *general* con el mismo número de variables enteras. Por fortuna, empezando con las ideas que se describirán en secciones subsecuentes, en la actualidad los mejores algoritmos de PE son ampliamente superiores al método de la numeración exhaustiva. La mejora en las dos décadas pasadas ha sido sorprendente. Los problemas PEB que requerían años de procesamiento para resolverse hace 20 años, en la actualidad pueden resolverse en segundos mediante el uso de paquetes de software comerciales (como, por ejemplo, el paquete CPLEX). Este aumento dramático en la velocidad se debe al enorme progreso en tres áreas: mejoras significativas en los algoritmos PEB (así como en otros algoritmos de PE), mejoras sorprendentes en algoritmos de programación lineal muy utilizados en los algoritmos de programación entera y el notable incremento de la velocidad de las computadoras (entre las cuales se incluyen las de escritorio). Como consecuencia de todo ello,

[3] De manera estricta, un problema de cobertura de conjunto no incluye *otras* restricciones funcionales como la última del ejemplo de programación de tripulaciones. Algunas veces se supone que todos los coeficientes de la función objetivo que se quiere minimizar son iguales a *uno*, y entonces cuando este supuesto no se cumple se usa el nombre de *problema de cobertura de conjunto ponderada*.

Recuadro de aplicación

Taco Bell Corporation cuenta con más de 6 500 restaurantes de servicio rápido en Estados Unidos y un mercado internacional que cada día crece más. Sirve alrededor de 2 000 millones de alimentos anualmente, lo cual genera aproximadamente 5.4 miles de millones de dólares de ingreso de ventas al año.

En cada restaurante Taco Bell, la cantidad de servicios es muy variable a lo largo del día (y de un día a otro), con una alta concentración durante las horas de comida. Por tanto, la determinación de cuántos empleados deben programarse para llevar a cabo qué funciones en el restaurante en todo momento, representa un problema complejo y tedioso.

Para enfrentar este problema, la gerencia de Taco Bell giró órdenes a un equipo de investigación de operaciones (que incluía a varios consultores) para que desarrollaran un nuevo sistema de administración del trabajo. El equipo concluyó que el sistema necesitaba tres componentes principales: 1) un *modelo de pronóstico* para predecir las transacciones de los clientes en cualquier momento, 2) un *modelo de simulación* (como los que se describien en el capítulo 20) para traducir las transacciones de los clientes en necesidades de trabajo y, 3) un *modelo de programación entera* con el fin de programar a los empleados para satisfacer las necesidades de trabajo y minimizar la nómina.

Las variables de decisión enteras de este modelo de programación entera para cualquier restaurante son el número de empleados asignados en cada turno que inicien su horario de trabajo en determinados tiempos específicos. La duración de estos turnos son también variables de decisión (restringidas a estar entre la duracion máxima y mínima permisible de un turno), pero en este caso variables de decisión *continuas*, por lo que el modelo es un modelo IP mezclado. Las restricciones principales especifican que el número de empleados que trabajan durante cada intervalo de 15 minutos debe ser mayor o igual al número mínimo que se requiere durante ese intervalo (de acuerdo con el modelo de pronóstico).

Este modelo PEM es parecido al *modelo de programación lineal* para asignar empleados a los turnos en el problema de United Airlines que se describió en el recuadro de aplicación de la sección 3.4. Sin embargo, la diferencia fundamental estriba en que el número de empleados que trabaja en turnos en los restaurantes Taco Bell es mucho menor que el de United Airlines, por lo que es necesario restringir dichas variables de decisión a valores enteros en el caso del modelo de taco Bell (mientras que los valores no enteros de una solución en el caso de United Airlines puede redondearse con facilidad a valores enteros con tan sólo una pequeña pérdida de precisión).

La implantación de este modelo MIP junto con los demás componentes del sistema de administración del trabajo ha brindado a Taco Bell *ahorros documentados* de **13 millones** de dólares *al año* en costos laborales.

Fuente: J. Hueter y W. Swart: "An Integrated Labor-Management System for taco Bell", en *Interfaces*, **28**(1): 75-91, enero-febrero de 1998. (En el sitio en internet de este libro, www.mhhe.com/hillier, se proporciona una liga hacia este artículo.)

una gran cantidad de problemas PEB de gran tamaño están siendo resueltos en la actualidad, lo que hubiera sido imposible en las décadas pasadas. Los mejores algoritmos de hoy en día son capaces de resolver *algunos* problemas PEB puros con cientos de miles de variables. Sin embargo, debido al *crecimiento exponencial*, ni los mejores algoritmos garantizan la solución de todos los problemas relativamente pequeños (con menos de 100 variables binarias o enteras). Según sus características, ciertos problemas pequeños pueden ser más difíciles de resolver que algunos más grandes.

Cuando se trata con variables enteras generales en lugar de variables binarias, el tamaño de los problemas que pueden resolverse tiende a ser mucho más pequeño. Sin embargo, hay excepciones. Por ejemplo, hace varios años, la versión profesional de CPLEX 8.0 resolvió de manera exitosa un problema de PE con 215 mil variables enteras generales, 75 mil restricciones funcionales y 6 millones de coeficientes de restricción distintos de cero. Con todo, las versiones más actuales de CPLEX son mucho más poderosas.

La segunda falacia es que si se eliminan algunas soluciones factibles (las no enteras) de un problema de programación lineal, será más fácil resolverlo. Por el contrario, sólo cuando todas estas soluciones factibles están ahí, se puede garantizar (vea la sección 5.1) que existe una solución factible en el vértice (FEV) [y por ende la solución básica factible (BF) correspondiente] que es óptima para el problema completo. Esta garantía es la clave de la extraordinaria eficiencia del método símplex. Como resultado, en general es mucho más sencillo resolver los problemas de programación lineal que los de programación entera.

Es lógico, entonces, que la mayor parte de los algoritmos exitosos de programación entera incorporen un algoritmo de programación lineal, como el método símplex (o el método símplex dual) lo más que puedan, y que relacionen partes del problema de PE bajo consideración con el problema correspondiente de programación lineal (es decir, el mismo problema con la restricción de valores enteros eliminada). En cualquier problema dado de PE, el problema correspondiente de programación lineal se conoce como su **relajamiento de PL**. El algoritmo que se presenta en las dos secciones siguientes ilustra cómo se puede usar una sucesión de relajamientos de PL para

porciones de un problema de programación entera, con objeto de resolver de manera eficaz un problema completo de PE.

Existe una situación especial en la que no es más difícil resolver el problema de PE que resolver una vez su relajamiento de PL al aplicar el método símplex; en este caso, la solución del relajamiento de PL satisface la restricción de valores enteros del problema de PE. Cuando ocurre esto, la solución también *debe* ser óptima para el problema de PE, puesto que se trata de la mejor solución entre todas las soluciones factibles del relajamiento de PL, que incluyen todas las soluciones factibles del problema de PE. Entonces, es normal que un algoritmo de programación entera comience con la aplicación del método símplex a su relajamiento de PL, para verificar si tiene lugar este acontecimiento fortuito.

Aunque, sin duda, casi siempre es accidental que la solución óptima del relajamiento de PL sea entera, en realidad existen varios *tipos especiales* de problemas de PE para los que este resultado se puede *garantizar*. Ya se vieron dos de estos casos especiales en los capítulos 8 y 9; éstos son el *problema del flujo de costo mínimo* (con parámetros enteros) y sus casos especiales (que incluyen el *problema de transporte*, el *problema de trasbordo*, el *problema de asignación*, el *problema de la ruta más corta* y el *de flujo máximo*). Esta garantía se puede dar en razón de la *estructura especial* que poseen estos tipos particulares de problemas —por ejemplo, vea la tabla 8.6—, que asegura que toda solución factible es entera, como se estableció en la propiedad de soluciones enteras que se presentó en las secciones 8.1 y 9.6. Por ello, estos tipos especiales de problemas de PE se pueden manejar como problemas de programación lineal, puesto que se pueden resolver por completo al aplicar las versiones simplificadas del método símplex.

Aunque toda esta simplificación no es usual, con frecuencia los problemas de PE que surgen en la práctica tienen *alguna* estructura especial que se puede aprovechar para simplificarlos. (Los ejemplos 2 y 3 de la sección anterior caen en esta categoría debido a sus restricciones de *alternativas mutuamente excluyentes* o las de *decisiones contingentes* o las restricciones de *cobertura de un conjunto*.) A veces se pueden resolver con éxito versiones muy grandes de estos problemas. Cada vez son más importantes los algoritmos especiales diseñados específicamente para explotar ciertos tipos de estructuras especiales de programación entera.

En consecuencia, los tres factores determinantes de la *dificultad computacional* de un problema de PE son 1) el *número de variables enteras*, 2) si estas variables enteras son *binarias* o *generales* y 3) cualquier *estructura especial* del problema. Esta situación es opuesta a la de programación lineal, en donde el número de restricciones (funcionales) es mucho más importante que el número de variables. En programación entera, el número de restricciones tiene *alguna* importancia —en especial si se van a resolver los relajamientos de PL—, pero es estrictamente secundario en el caso de los otros tres factores. En realidad, existen casos en los que *aumentar* el número de restricciones *disminuye* el tiempo de cálculo, puesto que se reduce el número de soluciones factibles. En los problemas de PEM es el número de variables *enteras* y no el número *total* de variables lo que es importante, pues las variables *continuas* casi no tienen efecto sobre el esfuerzo computacional.

Debido a que, en general, es mucho más difícil resolver los problemas de PE que los de programación lineal, a veces es tentador usar el procedimiento aproximado y aplicar el método símplex al relajamiento de PL y después *redondear* los valores no enteros a enteros en la solución que se obtuvo. Este enfoque puede ser adecuado en algunas aplicaciones, en especial si los valores de las variables son tan grandes que el redondeo introduce un error muy pequeño; sin embargo, debe tenerse cuidado al poner en práctica este procedimiento, pues se corren dos riesgos.

Una desventaja es que una solución óptima de programación lineal *no necesariamente es factible* después de redondearla. Con frecuencia es difícil decidir en qué sentido redondear para conservar la factibilidad. Incluso, podría ser necesario cambiar el valor de algunas variables en una o más unidades después de redondear. Para ilustrar este método, considere el siguiente problema:

Maximizar $\quad Z = x_2,$

sujeta a

$$-x_1 + x_2 \leq \frac{1}{2}$$

$$x_1 + x_2 \leq 3\frac{1}{2}$$

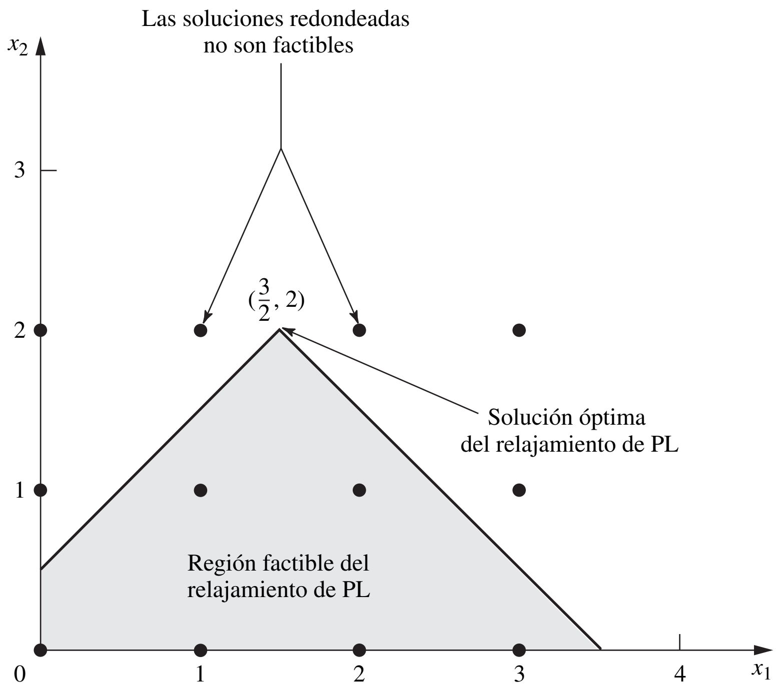


FIGURA 11.2
Ejemplo de un problema de PE donde la solución óptima del relajamiento de PL no se puede redondear de ninguna manera que conserve la factibilidad.

y

$$x_1 \geq 0, \qquad x_2 \geq 0$$
$$x_1, x_2 \text{ son enteros.}$$

Como se muestra en la figura 11.2, la solución óptima para el relajamiento de PL es $x_1 = 1\frac{1}{2}, x_2 = 2$, pero es imposible redondear la variable no entera x_1 a 1 o 2 (o a cualquier otro entero) y conservar la factibilidad. Esta factibilidad sólo se puede conservar si también se cambia el valor entero de x_2. Es fácil imaginar la complejidad a la que puede llevar este tipo de dificultades cuando se tiene en frente decenas o cientos de restricciones y variables.

Aun cuando una solución óptima para el relajamiento de PL se pueda redondear con éxito, todavía queda otra dificultad. No existe garantía de que esta solución redondeada sea la solución óptima de programación entera. En realidad, puede ser que se encuentre muy lejos del óptimo en términos del valor de la función objetivo. El siguiente problema ejemplifica este hecho:

$$\text{Maximizar} \qquad Z = x_1 + 5x_2,$$

sujeta a

$$x_1 + 10x_2 \leq 20$$
$$x_1 \qquad \leq 2$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0$$
$$x_1, x_2 \text{ son enteros.}$$

Como sólo se tienen dos variables de decisión, este problema se puede graficar como se muestra en la figura 11.3. Se puede usar esta gráfica o el método símplex para encontrar que la solución óptima del relajamiento de PL es $x_1 = 2, x_2 = \frac{9}{5}$, con $Z = 11$. Si no se dispusiera de una solución gráfica óptima —que sería el caso con más variables de decisión—, la variable con el valor no entero $x_2 = \frac{9}{5}$ se redondearía en la dirección factible a $x_2 = 1$. La solución entera que se obtendría sería $x_1 = 2, x_2 = 1$, de donde se obtiene $Z = 7$. Observe que esta solución está muy lejos de la óptima $(x_1, x_2) = (0, 2)$ donde $Z = 10$.

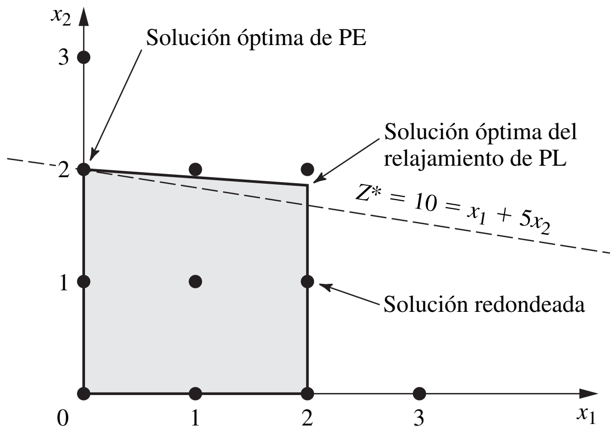


FIGURA 11.3
Ejemplo donde la solución del relajamiento de PL redondeada está muy lejos de ser óptima para el problema de PE.

Debido a estos dos riesgos, una mejor forma de tratar con problemas de programación entera grandes es usar uno de los *algoritmos heurísticos* disponibles. Estos algoritmos son muy eficientes para solucionar problemas grandes, pero no garantizan que se llegue a una solución óptima. Sin embargo, tienden a ser mucho más eficaces para encontrar excelentes soluciones factibles que el enfoque de redondeo que se acaba de analizar.

Uno de los desarrollos en particular emocionante de IO en años recientes ha sido el avance rápido en la elaboración de algoritmos heurísticos eficaces (llamados *metaheurísticos*) para diferentes problemas combinatorios como los de programación entera. Tres tipos sobresalientes de metaheurísticos —búsqueda de tabú, simulación de templado y algoritmos genéticos— se describirán en el capítulo 13. Estos complejos algoritmos metaheurísticos pueden incluso aplicarse a problemas de programación entera *no lineal* que tienen soluciones óptimas locales que es posible que estén lejos de la solución óptima global. También se pueden aplicar a diferentes problemas de *optimización combinatoria*, los cuales con frecuencia se pueden representar por un modelo que tiene variables enteras pero que también posee algunas restricciones que son más complicadas que las de un modelo de PE. (Tales aplicaciones se describirán más adelante en el capítulo 13.)

De regreso a la programación entera *lineal*, se dispone de un gran número de algoritmos para pequeños problemas de PE que los resuelven hasta llegar al óptimo. Sin embargo, ninguno posee una eficiencia computacional que se pueda siquiera comparar con el *método símplex* (excepto para los tipos especiales). Por tanto, el desarrollo de algoritmos de PE aún es un tema activo de investigación. Por fortuna, durante las dos últimas décadas se lograron algunos adelantos y se esperan más en los próximos años. Estos adelantos se verán en las secciones 11.8 y 11.9.

El algoritmo más popular de programación entera es la *técnica de ramificación y acotamiento* y las ideas relacionadas con la *enumeración implícita* de las soluciones factibles enteras, cuyos enfoques se analizarán más adelante. La siguiente sección presenta la técnica de ramificación y acotamiento en un contexto general y la ilustra con un algoritmo básico para enfrentar problemas de PEB. En la sección 11.7 se describe otro algoritmo del mismo tipo para problemas generales de PEM.

11.6 TÉCNICA DE RAMIFICACIÓN Y ACOTAMIENTO Y SUS APLICACIONES A LA PROGRAMACIÓN ENTERA BINARIA

Dado que cualquier problema acotado de programación *entera pura* tiene sólo un número finito de soluciones factibles, resulta natural considerar el uso de algún tipo de *procedimiento de enumeración* para encontrar una solución óptima. Desafortunadamente, como se mencionó en la sección anterior, este número finito puede ser, y casi siempre es, muy grande, por lo que es imperativo que cualquier procedimiento de enumeración se estructure con habilidad para que sólo sea necesario examinar una pequeña fracción de estas soluciones factibles. Por ejemplo, la programación dinámica (vea el capítulo 10) proporciona un procedimiento de este tipo para muchos problemas que tienen un número finito de soluciones factibles (aunque no es especialmente eficiente para la mayoría de los problemas de PE). Otro enfoque de este tipo lo proporciona la *técnica de ramifi-*

cación y acotamiento. Esta técnica y algunas de sus variantes se han aplicado con cierto éxito a diversos problemas de investigación de operaciones, pero es más conocida por sus aplicaciones a los problemas de programación entera.

La idea básica en la que se apoya la técnica de ramificación y acotamiento es *divide y conquistarás*. Como es demasiado complicado resolver en forma directa el problema original "grande", se divide en subproblemas cada vez más pequeños hasta que éstos se puedan vencer. La división (*ramificación*) se hace mediante una partición del conjunto completo de soluciones factibles en subconjuntos más pequeños. En parte, la conquista (*sondeo*) se hace mediante el *acotamiento* de la mejor solución del subconjunto para después descartar los subconjuntos cuya cota indique que no es posible que contenga una solución óptima para el problema original.

A continuación se describirá cada uno de estos pasos básicos —ramificación, acotamiento y sondeo— y se ilustrarán a través de la aplicación de un algoritmo de ramificación y acotamiento al ejemplo prototípico —el problema de California Manufacturing Co.— que se presentó en la sección 11.1 y que se repite en ésta (con las restricciones numeradas para referencias posteriores).

Maximizar $Z = 9x_1 + 5x_2 + 6x_3 + 4x_4$,

sujeta a

$$\begin{array}{ll}
(1) & 6x_1 + 3x_2 + 5x_3 + 2x_4 \le 10 \\
(2) & x_3 + x_4 \le 1 \\
(3) & -x_1 + x_3 \le 0 \\
(4) & -x_2 + x_4 \le 0
\end{array}$$

y

(5) x_j es binaria, para $j = 1, 2, 3, 4$.

Ramificación

Cuando se manejan variables binarias, la forma más sencilla de partir el conjunto de soluciones factibles es fijar el valor de una variable (por ejemplo, x_1) en $x_1 = 0$ para un subconjunto y en $x_1 = 1$ para el otro. Al hacer esto en el ejemplo prototipo, el problema completo queda dividido en dos subproblemas más pequeños, como se presentan a continuación.

S*ubproblema 1:*
Se fija $x_1 = 0$, con lo que el subproblema que resulta es

Maximizar $Z = 5x_2 + 6x_3 + 4x_4$,

sujeta a

$$\begin{array}{ll}
(1) & 3x_2 + 5x_3 + 2x_4 \le 10 \\
(2) & x_3 + x_4 \le 1 \\
(3) & x_3 \le 0 \\
(4) & -x_2 + x_4 \le 0
\end{array}$$

(5) x_j es binaria, para $j = 2, 3, 4$.

Subproblema 2:
Se fija $x_1 = 1$, con lo que el subproblema que se obtiene es

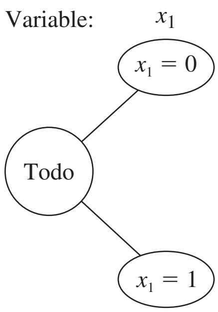
Maximizar $Z = 9 + 5x_2 + 6x_3 + 4x_4$,

sujeta a

$$\begin{array}{ll}
(1) & 3x_2 + 5x_3 + 2x_4 \le 4 \\
(2) & x_3 + x_4 \le 1 \\
(3) & x_3 \le 1 \\
(4) & -x_2 + x_4 \le 0
\end{array}$$

(5) x_j es binaria, para $j = 2, 3, 4$.

FIGURA 11.4
Árbol de solución creado por la ramificación de las dos primeras iteraciones del algoritmo de ramificación y acotamiento de PEB para el ejemplo de la sección 11.1.



En la figura 11.4 se muestra esto con una división (ramificación) en subproblemas, mediante un *árbol* (definido en la sección 9.2) con *ramas* (arcos) del nodo *Todo* —correspondiente al problema completo que contiene *todas* las soluciones factibles— a los dos nodos de dos subproblemas. Este árbol, cuyas ramas "crecerán" en cada iteración, se conoce como **árbol de soluciones** (o *árbol de enumeración*) del algoritmo. La variable que se utiliza en la ramificación de una iteración cuando se asignan valores a la variable (como x_1) se llama **variable de ramificación**. (Una parte importante de algunos algoritmos de ramificación y acotamiento son ciertos métodos elaborados para seleccionar la variable de ramificación pero, por simplicidad, en esta sección se seleccionan en su orden natural: $x_1, x_2, \ldots, x_n$.

Más adelante se verá que uno de estos subproblemas se puede vencer (sondear) de inmediato, mientras que el otro requiere una nueva división en subproblemas más pequeños si se establece $x_2 = 0$ o $x_2 = 1$.

En otros problemas de PE, donde las variables enteras pueden tener más de dos valores posibles, la ramificación se puede hacer a través de establecer la variable de ramificación igual a cada valor, lo que crea más de dos subproblemas. Otro buen enfoque es especificar el *intervalo* de valores (por ejemplo, $x_j \le 2$ o $x_j \ge 3$) para la variable de ramificación en cada nuevo subproblema. Éste es el enfoque que se usa en el algoritmo que se presenta en la sección 11.7.

Acotamiento

Ahora es necesario obtener, para cada subproblema, una *cota* que muestre el nivel de precisión de su mejor solución factible. La forma más común de hacerlo es resolver con rapidez un *relajamiento* sencillo del subproblema. Casi siempre, el **relajamiento** de un problema se obtiene *eliminando* ("relajando") un conjunto de restricciones que dificultan obtener una solución. En los problemas de PE, las restricciones más incómodas son las que requieren que las variables sean enteras. En consecuencia, el relajamiento que más se usa es el **relajamiento de PL** que elimina este conjunto de restricciones.

Para ilustrar con un ejemplo, considere primero el problema completo que se presentó en la sección 11.1. Su relajamiento de PL se obtiene mediante la sustitución del último renglón del modelo (x_j es binaria, para $j = 1, 2, 3, 4$) por la siguiente nueva (relajada) versión de su restricción (5)

$$(5) \qquad 0 \le x_j \le 1 \qquad \text{para } j = 1, 2, 3, 4.$$

Usando el método símplex para resolver con rapidez su relajación PL que se obtiene de su solución óptima

$$(x_1, x_2, x_3, x_4) = \left(\frac{5}{6}, 1, 0, 1\right), \qquad \text{con } Z = 16\frac{1}{2}.$$

Por tanto, $Z \le 16\frac{1}{2}$ para todas las soluciones factibles del problema original de PEB (puesto que estas soluciones son un subconjunto de las soluciones factibles del relajamiento de PL). En realidad, como se indica en el resumen del algoritmo, esta *cota* de $16\frac{1}{2}$ se puede redondear a 16, porque todos los coeficientes de la función objetivo son enteros y, por ende, deben dar un valor entero de Z.

Cota de todo el problema: $\qquad Z \le 16.$

Ahora se obtendrá, de la misma manera, las cotas de los subproblemas. En el subproblema 1, donde $x_1 = 0$, puede expresarse de manera apropiada en su relajación PL mediante la suma de la restricción que $x_1 \le 0$ puesto que cuando se combinan con la restricción actual $0 \le x_1 \le 1$ obliga a que $x_1 = 0$. De manera similar, si se fija $x_1 = 1$ en el subproblema 2 nos lleva a sumar la restricción $x_1 \ge 1$ para su relajamiento de PL. Aplicando el método símplex obtenemos las soluciones óptimas que se muestran en seguida de estos relajamientos de PL.

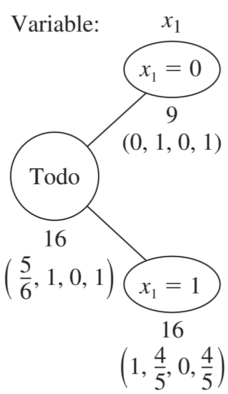
Relajamiento de PL del subproblema 1: (5) $\quad x_1 \le 0 \quad$ y $\quad 0 \le x_j \le 1 \qquad$ para $j = 1, 2, 3, 4$.
Solución óptima: $(x_1, x_2, x_3, x_4) = (0, 1, 0, 1)$ $\qquad$ con $Z = 9$.
Relajamiento de PL del subproblema 2: (5) $\quad x_1 \ge 1 \quad$ y $\quad 0 \le x_j \le 1 \qquad$ para $j = 1, 2, 3, 4$.
Solución óptima: $(x_1, x_2, x_3, x_4) = \left(1, \frac{4}{5}, 0, \frac{4}{5}\right)$ $\qquad$ con $Z = 16\frac{1}{5}$.

Las cotas que se obtienen son, entonces

Cota del subproblema 1: $Z \leq 9$,
Cota del subproblema 2: $Z \leq 16$.

En la figura 11.5 se resumen estos resultados. En ella, los números que se encuentran exactamente abajo de los nodos son las cotas, y debajo de cada cota se presenta la solución óptima que se obtuvo para el relajamiento de PL.

■ **FIGURA 11.5**
Árbol de solución después de la primera iteración del algoritmo de ramificación y acotamiento de PEB del ejemplo de la sección 11.1.



Sondeo

Un subproblema se puede conquistar (sondear), y, por tanto, ya no tomarse en cuenta, en las tres formas que se describen a continuación.

Una forma se ilustra con los resultados del subproblema 1 que se dieron en el nodo $x_1 = 0$, en la figura 11.5. Observe que la solución óptima (única) de este relajamiento de PL, $(x_1, x_2, x_3, x_4) = (0, 1, 0, 1)$, es una solución *entera*. En consecuencia, esta solución debe ser también la solución óptima del subproblema 1, que debe guardarse como la primera solución **incumbente** o de apoyo (la primera solución factible que se ha encontrado hasta ahora) del problema completo, junto con su valor de Z. Este valor se denota por

$Z^* =$ valor de Z de la solución de apoyo actual,

de manera que en este punto $Z^* = 9$. Como esta solución se guarda, no hay razón para seguir considerando el subproblema 1 y ramificar el nodo $x_1 = 0$. Hacerlo sólo conduciría a otras soluciones factibles inferiores a la de apoyo que no son de interés. Ahora, como ya se resolvió, el subproblema 1 se **sondea** (elimina).

Los resultados anteriores sugieren una segunda prueba de sondeo importante. Como $Z^* = 9$, no existe razón alguna para tomar en cuenta ningún subproblema cuya *cota* sea ≤ 9, puesto que tales subproblemas no pueden tener soluciones factibles mejores que la *incumbente*. Establecido en forma general, un subproblema se sondea siempre que su

Cota $\leq Z^*$.

Este resultado no ocurre en la iteración actual del ejemplo porque el subproblema 2 tiene una cota de 16 que es mayor que 9. No obstante, puede ocurrir más adelante, para los **descendientes** de este subproblema (nuevos problemas más pequeños creados al ramificar más este subproblema y quizá después al ramificarlo más en "generaciones" subsecuentes). Lo que es más, a medida que se encuentren nuevas soluciones de apoyo con valores más grandes de Z^* será más fácil *sondear* de esta manera.

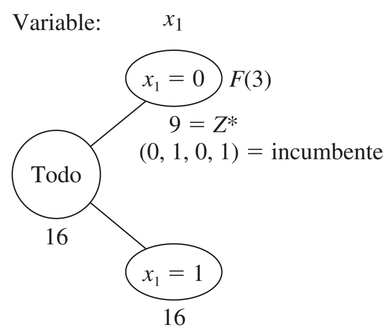
La tercera forma de sondeo es bastante directa. Si el método símplex encuentra que el relajamiento de PL de un subproblema *no tiene soluciones factibles*, entonces el subproblema en sí *no debe tener soluciones factibles*, de forma que puede eliminarse (sondearse).

En los tres casos, se lleva a cabo la búsqueda de una solución óptima mediante la investigación de sólo aquellos subproblemas que es posible que tengan una mejor solución factible que la de apoyo actual.

Resumen de las pruebas de sondeo. Un subproblema se *sondea* (se descarta para su consideración posterior) si

Prueba 1: Su cota $\leq Z^*$,
o
Prueba 2: Su relajamiento de PL no tiene soluciones factibles,
o
Prueba 3: La solución óptima para su relajamiento de PL es *entera*. (Si esta solución es mejor que la de apoyo, se convierte en la nueva solución de apoyo y se aplica de nuevo la prueba 1 a todos los subproblemas no sondeados, con la nueva mejor Z^*.)

En la figura 11.6 se resumen los resultados después de aplicar estas tres pruebas a los subproblemas 1 y 2; se muestra el *árbol de solución* actual. Sólo el subproblema 1 fue sondeado por la prueba 3, como lo indica $F(3)$ junto al nodo $x_1 = 0$. La solución de apoyo también se identifica debajo de este nodo.



■ **FIGURA 11.6**
Árbol de solución después de la primera iteración del algoritmo de ramificación y acotamiento de PEB para el ejemplo de la sección 11.1.

Las iteraciones siguientes ilustrarán la aplicación exitosa de las tres pruebas, pero antes de continuar con el ejemplo se resumirá el algoritmo que se aplica a este problema de PEB. (El algoritmo supone que la función objetivo se *maximizará*, que todos los coeficientes de la función objetivo son enteros y que, para efectos de simplicidad, el orden de las variables de la ramificación es $x_1, x_2, \ldots, x_n$. Como se pudo observar anteriormente, gran parte de los algoritmos de ramificación y acotamiento utilizan complejos métodos para seleccionar las variables de ramificación.)

Resumen de la técnica de ramificación y acotamiento de PEB

Paso inicial: Se establece $Z^* = -\infty$. Se aplica al problema completo el paso de acotamiento, el de sondeo y la prueba de optimalidad que se describió. Si no se sondea, este problema se clasifica como el único "subproblema" restante para realizar la primera iteración completa.

Pasos de cada iteración:

1. *Ramificación:* Entre los subproblemas *restantes* (no sondeados) se elige el de creación *más reciente*. (Los empates se rompen con el que tenga la *cota más grande*.) Se ramifica este nodo de ese subproblema para crear dos nuevos subproblemas luego de establecer la siguiente variable (la variable de ramificación), ya sea en 0 o en 1.
2. *Acotamiento:* En cada nuevo subproblema, aplique el método símplex a su relajación LP para obtener una solución óptima que incluya el valor de Z para esta relajación LP. Si este valor de Z no es entero, redondéelo a un entero. (Si lo era, no es necesario ningún cambio.) Este valor entero de Z es el *acotado* para el subproblema.
3. *Sondeo:* A cada nuevo subproblema se le aplican las tres pruebas de sondeo que se resumieron antes y se descartan aquellos que son sondeados por cualquiera de las tres pruebas.

Prueba de optimalidad: El proceso termina cuando no existen subproblemas *restantes*; la solución de *apoyo* actual es óptima.[4] De otra manera, se realiza otra iteración.

El paso de ramificación de este algoritmo merece un comentario respecto de la razón por la que se elige de esta manera el subproblema que se va a ramificar. Una opción que no se consideró sería elegir el subproblema restante con la *mejor cota*, puesto que este subproblema sería el más prometedor en cuanto a que puede contener la solución óptima del problema completo. La razón para usar la opción de seleccionar el subproblema de *creación más reciente* es que los *relajamientos* de PL se resuelven en el paso de acotamiento. En lugar de iniciar el método símplex cada vez desde el principio, cada relajamiento de PL en los problemas de gran escala se puede resolver casi siempre por *reoptimización*. Esta reoptimización implica revisar la tabla símplex final de relajamiento de PL anterior, según sea necesario, debido a las pocas diferencias en el modelo (igual que para el análisis de sensibilidad), y después aplicar unas cuantas iteraciones, quizá del método símplex dual. La reoptimización tiende a ser *mucho* más rápida que comenzar desde el principio, *siempre que* el

[4] Si no existe una solución de apoyo o incumbente, la conclusión es que el problema no tiene soluciones factibles.

modelo anterior y el actual sean parecidos. Los modelos tenderán a parecerse si se usa la regla de ramificación descrita, pero *no* cuando van de un lado a otro en el árbol de solución, como cuando se elige el subproblema con la mejor cota.

Terminación del ejemplo

El patrón de las iteraciones que faltan será muy similar al de la primera iteración que se acaba de describir, excepto por la forma en que se realiza el sondeo. Por tanto, se resumirán los pasos de ramificación y acotamiento y se hará hincapié en el paso de sondeo.

Iteración 2. El único subproblema que queda corresponde al nodo $x_1 = 1$ en la figura 11.6, por lo que se ramificará este nodo para crear dos nuevos subproblemas.

Subproblema 3:
Se fija $x_1 = 1, x_2 = 0$ de manera que el subproblema que resulta es

$$\text{Maximizar} \qquad Z = 9 + 6x_3 + 4x_4,$$

sujeta a

$$\begin{array}{ll} (1) & 5x_3 + 2x_4 \leq 4 \\ (2) & x_3 + x_4 \leq 1 \\ (3) & x_3 \leq 1 \\ (4) & x_4 \leq 0 \\ (5) & x_j \text{ es binaria}, \qquad \text{para } j = 3, 4. \end{array}$$

Subproblema 4:
Se fija $x_1 = 1, x_2 = 1$ de manera que el subproblema que resulta es

$$\text{Maximizar} \qquad Z = 14 + 6x_3 + 4x_4,$$

sujeta a

$$\begin{array}{ll} (1) & 5x_3 + 2x_4 \leq 1 \\ (2) & x_3 + x_4 \leq 1 \\ (3) & x_3 \leq 1 \\ (4) & x_4 \leq 1 \\ (5) & x_j \text{ es binaria}, \qquad \text{para } j = 3, 4. \end{array}$$

Los relajamientos de PL de estos subproblemas se obtienen consumando la restricción adicional que se muestra a continuación de la versión relajada de la restricción (5). Sus soluciones óptimas se muestran también a continuación.

Relajamiento de PL del subproblema 3: (5) $x_1 \geq 1, \quad x_2 \leq 0 \quad \text{y} \quad 0 \leq x_j \leq 1$ para $j = 1, 2, 3, 4$.

Solución óptima: $(x_1, x_2, x_3, x_4) = \left(1, 0, \frac{4}{5}, 0\right)$ con $Z = 13\frac{4}{5}$,

Relajamiento de PL del subproblema 4: (5) $x_1 \geq 1, \quad x_2 \geq 1 \quad \text{y} \quad 0 \leq x_j \leq 1$ para $j = 1, 2, 3, 4$.

Solución óptima: $(x_1, x_2, x_3, x_4) = \left(1, 1, 0, \frac{1}{2}\right)$ con $Z = 16$.

Las cotas que resultan para los subproblemas son

Cota del subproblema 3: $Z \leq 13$,
Cota del subproblema 4: $Z \leq 16$.

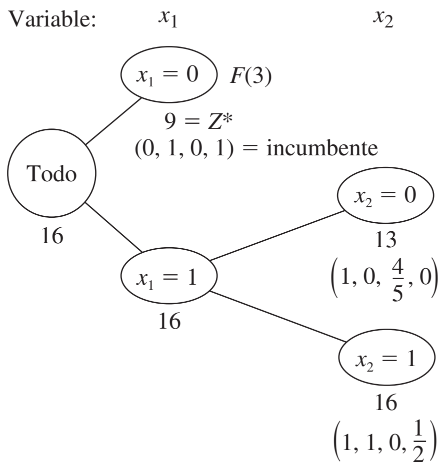


FIGURA 11.7
Árbol de solución después de la iteración 2 del algoritmo de ramificación y acotamiento de PEB del ejemplo de la sección 11.1.

Observe que estas cotas son más grandes que $Z^* = 9$, por lo que la prueba de sondeo 1 fracasa en ambos casos. La prueba 2 también fracasa, puesto que ambos relajamientos de PL tienen soluciones factibles (como lo indica la existencia de una solución óptima). De la misma manera, la prueba 3 fracasa porque ambas soluciones óptimas incluyen variables con valores no enteros.

En la figura 11.7 se muestra el árbol de solución resultante en este punto. La falta de una F a la derecha de cualquiera de los nuevos nodos indica que ambos quedan sin sondear.

Iteración 3. Hasta aquí, el algoritmo creó cuatro subproblemas. El subproblema 1 fue sondeado y el subproblema 2 se sustituyó por (se dividió en) los subproblemas 3 y 4, pero estos dos quedan bajo consideración. Como fueron creados al mismo tiempo y el subproblema 4 ($x_1 = 1$, $x_2 = 1$) tiene una *cota* más grande ($16 > 13$), la siguiente ramificación se hace desde el nodo $(x_1, x_2) = (1, 1)$ en el árbol de solución, lo que crea los siguientes subproblemas nuevos (en donde la restricción 3 desaparece porque no contiene a x_4).

Subproblema 5:
Se fija $x_1 = 1$, $x_2 = 1$, $x_3 = 0$ y el subproblema que resulta es

Maximizar $\quad Z = 14 + 4x_4,$

sujeta a

(1) $\quad 2x_4 \leq 1$
(2), (4) $\quad x_4 \leq 1 \quad$ (dos veces)
(5) $\quad x_4$ es binaria.

Subproblema 6:
Se fija $x_1 = 1$, $x_2 = 1$, $x_3 = 1$ de manera que el subproblema que resulta es

Maximizar $\quad Z = 20 + 4x_4,$

sujeta a

(1) $\quad 2x_4 \leq -4$
(2) $\quad x_4 \leq 0$
(4) $\quad x_4 \leq 1$
(5) $\quad x_4$ es binaria.

Las relajaciones PL correspondientes tienen la versión relajada de la restricción (5), la solución óptima y la acotada (cuando existe) que se muestra en seguida.

Relajamiento de PL del subproblema 5:

$$(5) \quad x_1 \geq 1, \quad x_2 \geq 1, \quad x_3 \leq 0 \quad \text{y} \quad 0 \leq x_j \leq 1 \quad \text{para } j = 1, 2, 3, 4.$$

Solución óptima: $(x_1, x_2, x_3, x_4) = \left(1, 1, 0, \frac{1}{2}\right)$, con $Z = 16$.

Cota: $Z \leq 16$

Relajamiento de PL del subproblema 6:

$$(5) \quad x_1 \geq 1, \quad x_2 \geq 1, \quad x_3 \geq 1 \quad \text{y} \quad 0 \leq x_j \leq 1 \quad \text{para } j = 1, 2, 3, 4.$$

Solucion óptima: Ninguna, puesto que no hay soluciones factibles.

Cota: ninguna

En ambos subproblemas, la versión relajada de la restricción (5) tiene el efecto de fijar los valores de x_1, x_2 y x_3 en los deseados y, por tanto, requieren que $0 \leq x_4 \leq 1$. Por ello, los relajamientos PL de estos subproblemas se reducen a las premisas de los subproblemas que se plantearon antes, excepto por el reemplazo de la restricción (5) por $0 \leq x_4 \leq 1$. Si se reducen estas relajaciones PL a problemas de una variable (más los valores fijos de x_1, x_2 y x_3) se hace más sencillo ver que la solución óptima de la relajación PL del subproblema 5 es, en realidad, el que se proporcionó antes. De manera similar, observe que la combinación de la restricción 1 y $0 \leq x_4 \leq 1$ en el relajamiento de PL del subproblema 6 evita las soluciones factibles. Por tanto, este subproblema se sondea mediante la prueba 2. No obstante, el subproblema 5 no pasa esta prueba, ni la 1 ($16 > 9$), ni la prueba 3 ($x_4 = \frac{1}{2}$ no es un entero), así que permanece bajo consideración.

En la figura 11.8 se muestra el árbol de solución.

Iteración 4. Los subproblemas que corresponden a los nodos (1, 0) y (1, 1, 0) de la figura 11.8 permanecen bajo consideración, pero el último nodo es el de creación más reciente, por lo que se selecciona para ramificación. Como la variable de ramificación x_4 es la *última* variable, si se fija su valor en 0 o en 1, se crea *una solución* y no un subproblema que requiera más investigación. Las soluciones creadas son

$x_4 = 0$: $(x_1, x_2, x_3, x_4) = (1, 1, 0, 0)$ es factible, con $Z = 14$,
$x_4 = 1$: $(x_1, x_2, x_3, x_4) = (1, 1, 0, 1)$ es factible.

Mediante la aplicación formal de las pruebas de sondeo, se comprueba que la primera solución pasa la prueba 3 y la segunda la prueba 2. Además, la primera solución factible es mejor que la de apoyo ($14 > 9$), por lo cual se convierte en la nueva solución de apoyo, con $Z^* = 14$.

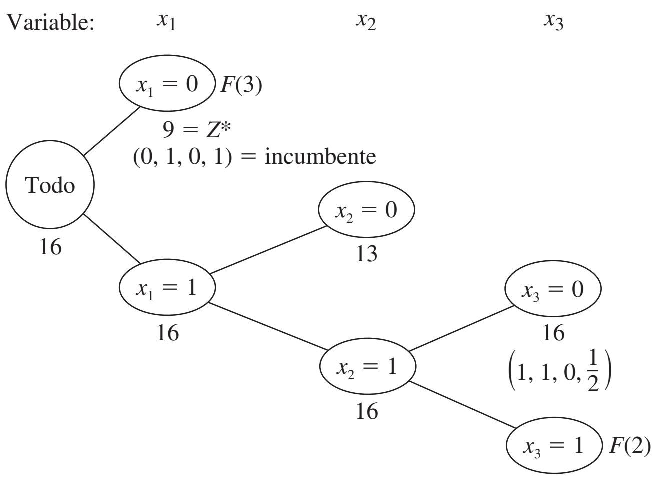


FIGURA 11.8 Árbol de solución después de la iteración 3 del algoritmo de ramificación y acotamiento de PEB del ejemplo de la sección 11.1.

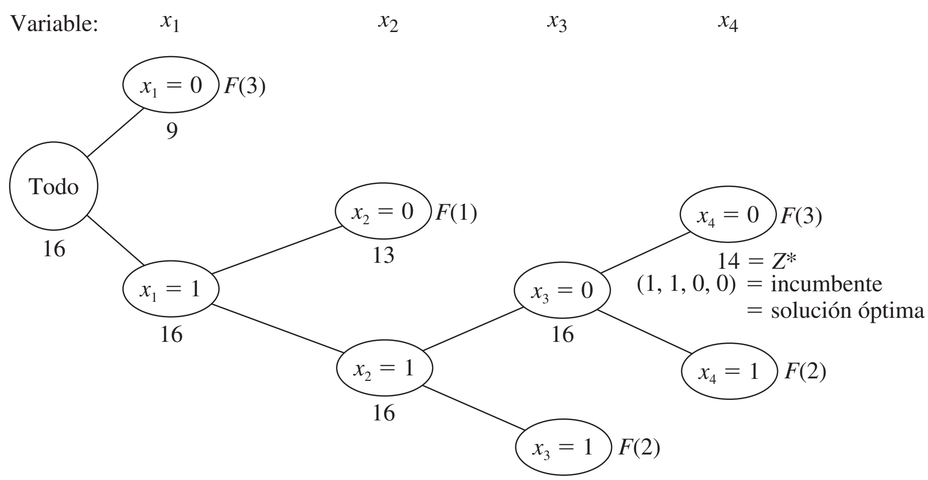


FIGURA 11.9
Árbol de solución después de la última (cuarta) iteración del algoritmo de ramificación y acotamiento de PEB del ejemplo de la sección 11.1.

Como se encontró una nueva solución de apoyo, se vuelve a aplicar la prueba 1 con el nuevo valor más grande de Z^* al único subproblema que queda, el del nodo (1, 0).

Subproblema 3:

$$\text{Cota} = 13 \leq Z^* = 14.$$

Entonces, este subproblema queda sondeado.

Se tiene ahora el árbol de solución de la figura 11.9. Observe que *no* hay subproblemas *restantes* (sin sondear). En consecuencia, la prueba de optimalidad indica que la solución de apoyo actual

$$(x_1, x_2, x_3, x_4) = (1, 1, 0, 0)$$

es óptima, y el problema termina aquí.

El OR Tutor incluye **otro ejemplo** de la aplicación de este algoritmo. También una rutina interactiva para ejecutar el algoritmo. Como es usual, los archivos de Excel, LINGO/LINDO y MPL/CPLEX para este capítulo muestran cómo se aplican las versiones para estudiantes de estos paquetes a los ejemplos del capítulo. Los algoritmos que se usan para solucionar los problemas de PEB son similares a los que se describen aquí.[5]

Otras opciones con la técnica de ramificación y acotamiento

En esta sección se ha ilustrado la técnica de ramificación y acotamiento con la descripción de un algoritmo básico para resolver problemas de programación entera binaria. Sin embargo, el marco general de trabajo de la técnica de ramificación y acotamiento proporciona una gran flexibilidad para diseñar un algoritmo específico para cualquier tipo de problema. Se dispone de muchas opciones y el diseño de un algoritmo eficiente requiere una adaptación para que se ajuste a la estructura particular del tipo de problema.

Cada algoritmo de ramificación y acotamiento contiene los mismos pasos básicos de *ramificación*, *acotamiento* y *sondeo*. La flexibilidad consiste en cómo se realizan estos pasos.

La *ramificación* siempre implica *seleccionar* un subproblema restante y *dividirlo* en subproblemas más pequeños. En este caso, la flexibilidad se encuentra en las reglas para seleccionar y dividir. El algoritmo seleccionado elige el subproblema de *creación más reciente*, porque este método es muy eficiente para *reoptimizar* cada relajamiento de PL a partir de la anterior. La otra regla popular es elegir el subproblema con la *mejor cota* puesto que tiende a llegar más rápido a mejores

[5] En la versión profesional de LINGO, LINDO y CPLEX, el algoritmo de PEB utiliza también una variedad de técnicas elaboradas similares a las descritas en la sección 11.8.

soluciones de apoyo y más sondeos. También se puede usar una combinación de estas dos reglas. La *división*, por lo general (pero no siempre) se hace seleccionando una *variable de ramificación* a la cual se asignan ya sea valores individuales (como en el algoritmo de PEB), o intervalos de valores (como en el algoritmo de la siguiente sección). Algunos algoritmos más elaborados utilizan una regla para elegir de manera estratégica una variable de ramificación que conduzca a sondeos rápidos. Este enfoque es mucho más eficiente que la regla que utiliza nuestro algoritmo BP de seleccionar simplemente las variables de ramificación en su orden natural, $x_1, x_2, \ldots, x_n$. Por ejemplo, una desventaja importante de esta simple regla para seleccionar la variable de ramificación es que si la variable tiene un valor entero en la solución óptima para la relajación PL del subproblema que se está ramificando, el subproblema siguiente que fija esta variable a este mismo valor entero también tendrá la misma solución óptima para su relajación PL, por lo que no se habrá avanzado nada hacia el sondeo. Por tanto, las opciones más estratégicas para seleccionar la variable de ramificación deben hacer algo parecido a seleccionar la variable cuyo valor en la solución óptima para la relajación PL del subproblema actual esté *más lejos* de ser un entero.

El *acotamiento* casi siempre se hace mediante la resolución de un *relajamiento*. Sin embargo, existen muchas formas de ellos. Por ejemplo, considere el **relajamiento de Lagrange**, en donde se *elimina* el conjunto completo de restricciones funcionales, en notación matricial $\mathbf{Ax} \leq \mathbf{b}$ (excepto quizá algunas restricciones "convenientes"), y después la función objetivo

$$\text{Maximizar} \qquad Z = \mathbf{cx},$$

se reemplaza por

$$\text{Maximizar} \qquad Z_R = \mathbf{cx} - \boldsymbol{\lambda}(\mathbf{Ax} - \mathbf{b}),$$

donde el vector fijo es $\boldsymbol{\lambda} \geq \mathbf{0}$. Si $\mathbf{x}^*$ es una solución óptima para el problema original, su $Z \leq Z_R$, por lo que cuando se obtiene un valor óptimo de Z_R con el relajamiento de Lagrange, se llega a una *cota* válida. Si $\boldsymbol{\lambda}$ se elige bien, esta cota tiende a ser razonablemente cerrada (al menos comparable con la cota del relajamiento de PL). Sin restricciones funcionales, este relajamiento se puede resolver muy rápido. Las desventajas son que las pruebas de sondeo 2 y 3 (revisadas) no son tan poderosas como las del relajamiento de PL.

En términos generales, las dos características que se buscan al elegir un relajamiento son que se pueda resolver relativamente rápido y que proporcione una cota bastante cerrada. Ninguna de las dos características por sí sola es adecuada. El relajamiento de PL se usa más porque proporciona un excelente trueque entre estos dos factores.

Una opción que se emplea en ciertas ocasiones es la solución rápida de un relajamiento y después, si no se logra el sondeo, se cierra el relajamiento de alguna forma para obtener una mejor cota.

El *sondeo* casi siempre se hace como se describió en el algoritmo de PEB. Los tres criterios para sondear se pueden establecer en términos más generales como sigue.

Resumen de criterios de sondeo. Un subproblema se *sondea* si el análisis de su *relajamiento* revela que

Criterio 1: Las soluciones factibles del problema cumplen con que $Z \leq Z^*$, o
Criterio 2: El subproblema no tiene soluciones factibles, o
Criterio 3: Se encontró una solución óptima del subproblema.

Igual que en el caso del algoritmo de PEB, los dos primeros criterios se aplican cuando se resuelve un relajamiento para obtener una cota sobre el subproblema y después verificar si esta cota es $\leq Z^*$ (prueba 1) o si el relajamiento no tiene soluciones factibles (prueba 2). Si el relajamiento difiere del subproblema *sólo* en que se eliminaron (o se relajaron) algunas restricciones, se aplica el tercer criterio para verificar si la solución óptima del relajamiento es *factible* para el subproblema, en cuyo caso debe ser *óptima* para él. En otros relajamientos (como el de Lagrange), se requiere un análisis adicional para determinar si la solución óptima para el relajamiento es óptima para el subproblema.

Si el problema original involucra *minimización* en lugar de maximización, se tienen dos opciones. Una es convertirlo a maximización en la forma usual (vea la sección 4.6). La otra es con-

vertir directamente el algoritmo de ramificación y acotamiento a la forma de minimización, lo que requiere cambiar la dirección de la desigualdad de la prueba de sondeo 1, de

¿La cota del subproblema es $\leq Z^*$?

a

¿La cota del subproblema es $\geq Z^*$?

Cuando se usa esta última desigualdad, si el valor de Z de la solución óptima para la relajación LP del subproblema no es entero, se redondea a un entero con el fin de obtener la cota del subproblema.

Hasta ahora se ha descrito cómo usar la técnica de ramificación y acotamiento para encontrar *una* solución óptima. Sin embargo, en el caso de empates a veces es deseable identificar *todas* estas soluciones óptimas para que la elección final entre ellas pueda hacerse sobre factores intangibles que no se incorporaron al modelo matemático. Para encontrar todas es necesario introducir unos cuantos cambios al procedimiento. Primero, se cambia la desigualdad débil en la prueba de sondeo 1 (¿La cota del subproblema es $\leq Z^*$?) por una desigualdad estricta (¿La cota del subproblema es $< Z^*$?) para que el sondeo no ocurra si el subproblema puede tener una solución factible *igual* a la de apoyo. Segundo, si la prueba de sondeo 3 pasa y la solución óptima para el subproblema tiene $Z = Z^*$, se almacena como *otra* solución de apoyo (empatada). Tercero, si la prueba de sondeo 3 proporciona una nueva solución de apoyo (empatada o no), se verifica si la solución óptima que se obtuvo para el *relajamiento* es *única*. Si no lo es, se identifican las otras soluciones óptimas del relajamiento y se verifica que sean óptimas también para el subproblema, en cuyo caso se convierten en soluciones de apoyo. Por último, cuando la *prueba de optimalidad* encuentra que no hay subconjuntos *restantes* (sin sondear), *todas* las soluciones de *apoyo* actuales serán soluciones *óptimas*.

Para terminar, debe hacerse notar que más que encontrar una solución óptima, la técnica de ramificación y acotamiento se puede usar también para encontrar una solución *cercana a la óptima*, en general, con mucho menos esfuerzo computacional. En algunas aplicaciones, una solución es "bastante buena" si su Z está lo "suficientemente cerca" del valor de Z de la solución óptima (llámese Z^{**}). *Suficientemente cerca* se puede definir como

$$Z^{**} - K \leq Z \qquad \text{o como} \qquad (1 - \alpha)Z^{**} \leq Z$$

para una constante (positiva) específica K o α. Por ejemplo, si se elige la segunda definición y $\alpha = 0.05$, se requiere que la solución se encuentre a 5% o menos de la óptima. En consecuencia, si se supiera que el valor de Z de la solución de apoyo actual (Z^*) satisface una de las dos desigualdades

$$Z^{**} - K \leq Z^* \qquad \text{o} \qquad (1 - \alpha)Z^{**} \leq Z^*$$

entonces el procedimiento podría terminar de inmediato con la elección de la solución de apoyo como la solución cercana a la óptima deseada. Aunque, de hecho, el procedimiento no identifica una solución óptima y la Z^{**} correspondiente, si esta solución (desconocida) es factible (y, por tanto, óptima) para el subproblema actual bajo investigación, entonces la prueba de sondeo 1 encuentra una cota superior tal que

$$Z^{**} \leq \text{cota}$$

de manera que

$$\text{Cota} - K \leq Z^* \qquad \text{o} \qquad (1 - \alpha)\text{cota} \leq Z^*$$

lo que implicaría que se satisface la desigualdad correspondiente en la afirmación anterior. Aun cuando esta solución no es factible para el subproblema actual, se obtiene una cota superior válida para el valor de Z de la solución óptima de ese subproblema. Así, es suficiente satisfacer cualquiera de las dos desigualdades para sondear este subproblema porque la solución de apoyo debe estar "bastante cerca" de la solución óptima del subproblema.

Por tanto, para encontrar una solución que esté bastante cerca de la óptima, sólo se necesita un cambio en el algoritmo de ramificación y acotamiento. Este cambio es sustituir la prueba de sondeo 1 en el subproblema

$$\text{Cota} \leq Z^*?$$

por

$$\text{Cota} - K \leq Z^*?$$

o bien

$$(1 - \alpha)(\text{cota}) \leq Z^*?$$

y realizar esta prueba *después* de la prueba 3 (para que una solución factible que se encuentra con $Z > Z^*$ se conserve como la nueva solución de apoyo). La razón para que esta prueba 1 más débil sea suficiente es que no importa qué tan cerca de la cota del subproblema se encuentre Z de la solución óptima (desconocida) del subproblema, la solución de apoyo todavía estará bastante cerca de esta solución (si se cumple la nueva desigualdad) como para que no sea necesario seguir analizando el subproblema. Cuando no hay subproblemas restantes, la solución de apoyo actual será la solución *cercana a la óptima* que se desea. Sin embargo, es mucho más sencillo sondear con esta nueva prueba (en cualquiera de sus formas), por lo que el algoritmo trabajará más rápido. Si se trata de problemas muy grandes, esta aceleración puede significar la diferencia entre terminar con una solución que sea cercana a la óptima y nunca terminar. En el caso de problemas muy grandes que surgen en la práctica, como el modelo brinda sólo una representación idealizada del problema real, es suficiente para todo propósito en la práctica encontrar una solución casi óptima para el modelo de esta forma. Por ende, en la práctica, se utiliza este atajo muy frecuentemente.

11.7 ALGORITMO DE RAMIFICACIÓN Y ACOTAMIENTO PARA PROGRAMACIÓN ENTERA MIXTA

A continuación se estudiará el problema general de programación entera mixta, PEM, donde *algunas* variables (por ejemplo, I de ellas) están restringidas a valores enteros (pero no necesariamente sólo 0 y 1), y el resto son variables continuas comunes. Por conveniencia en la notación, estas variables se ordenarán de manera que las primeras I de ellas sean variables con *restricción de enteros*. La forma general del problema que se va a estudiar es

$$\text{Maximizar} \qquad Z = \sum_{j=1}^{n} c_j x_j,$$

sujeta a

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i, \qquad \text{para } i = 1, 2, \ldots, m,$$

y

$$x_j \geq 0, \qquad \text{para } j = 1, 2, \ldots, n,$$
$$x_j \text{ es entero}, \qquad \text{para } j = 1, 2, \ldots, I;\ I \leq n.$$

(Cuando $I = n$, este problema se convierte en uno de PE pura.)

Se describirá un algoritmo básico de ramificación y acotamiento para resolver este problema que, un poco más elaborado, ha proporcionado el enfoque estándar a la PEM. La estructura de este algoritmo fue desarrollada por R. J. Dakin,[6] con base en un algoritmo anterior que se debe a A. H. Land y A. G. Doig.[7]

La estructura de este algoritmo es bastante parecida a la del algoritmo de PEB que se presentó en la sección anterior. De nuevo la solución de un *relajamiento de PL* proporciona la base tanto para el paso de *acotamiento* como para el de *sondeo*. En realidad, sólo se necesitan cuatro cambios al algoritmo de PEB para manejar la generalización de variables enteras *binarias generales* y de PE *pura* a PE *mixta*.

---

[6] R. J. Dakin, "A Tree Search Algorithm for Mixed Integer Programming Problems", en *Computer Journal*, **8**(3): 250-255, 1965.

[7] A. H. Land y A. G. Doig, "An Automatic Method of Solving Discrete Programming Problems", en *Econometrica*, **28**: 497-520, 1960.

Un cambio involucra la elección de la *variable de ramificación*. Antes se elegía de manera automática la *siguiente* variable en el orden natural ($x_1, x_2, \ldots, x_n$). Ahora, las únicas variables que se toman en cuenta son las variables con *restricción de enteros* que tienen un valor *no entero* en la solución óptima del relajamiento de PL para el subproblema actual. La regla para elegir entre estas variables será seleccionar la *primera* en el orden natural. (Los paquetes de producción casi siempre utilizan una regla más complicada.)

El segundo cambio se refiere a los valores asignados a la variable de ramificación para crear los nuevos subproblemas. Antes, la variable *binaria* se fijaba en 0 y 1, respectivamente, para los dos nuevos subproblemas. Ahora, la variable *general* con restricción de entera puede tener un número grande de valores enteros posibles, y sería ineficiente crear *y* analizar *muchos* subproblemas mediante la determinación de cada valor entero de la variable. Entonces, se crean sólo *dos* nuevos subproblemas (como antes) al especificar *dos intervalos* de valores de la variable.

Para aclarar cómo se procede, sea x_j la variable de ramificación actual y x_j^* su valor (no entero) en la solución óptima del relajamiento de PL del subproblema actual. Se usan corchetes para denotar

$$[x_j^*] = \text{entero más grande} \leq x_j^*,$$

por lo cual, para los intervalos de valores de los dos nuevos subproblemas se tiene

$$x_j \leq [x_j^*] \qquad \text{y} \qquad x_j \geq [x_j^*] + 1,$$

respectivamente. Cada desigualdad se convierte en una *restricción adicional* del nuevo subproblema. Por ejemplo, si $x_j^* = 3\frac{1}{2}$, entonces

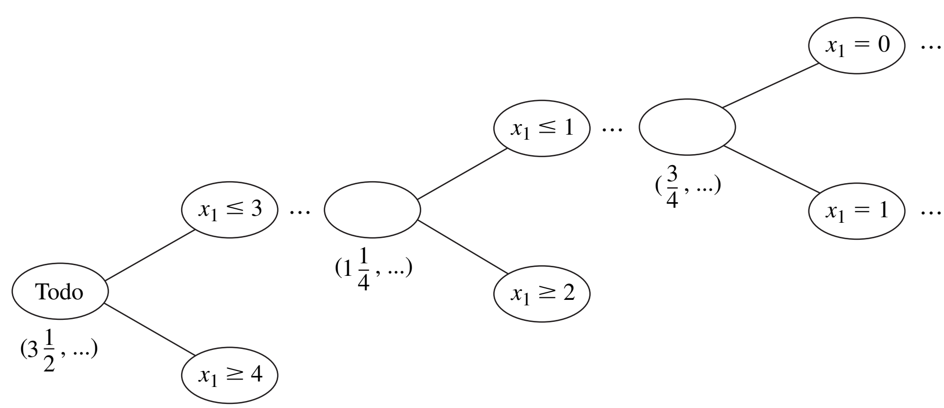
$$x_j \leq 3 \qquad \text{y} \qquad x_j \geq 4$$

son las restricciones adicionales respectivas de los dos nuevos subproblemas.

Cuando se combinan los dos cambios al algoritmo de PEB que se acaba de describir, puede ocurrir un interesante fenómeno de una *variable de ramificación recurrente*. Para ilustrarla, como se muestra en la figura 11.10, sea $j = 1$ en el ejemplo anterior donde $x_j^* = 3\frac{1}{2}$ y considere el nuevo subproblema con $x_1 \leq 3$. Cuando se resuelve el relajamiento de PL para un descendiente de este subproblema se obtiene $x_1^* = 1\frac{1}{4}$. Entonces, x_1 *recurre* como variable de ramificación y los dos nuevos subproblemas creados tienen la restricción adicional $x_1 \leq 1$ y $x_1 \geq 2$, respectivamente (así como la restricción adicional anterior, $x_1 \leq 3$). Más adelante, cuando se resuelva el relajamiento de PL de un descendiente de, por ejemplo, $x_1 \leq 1$, se supondrá que $x_1^* = \frac{3}{4}$. Entonces x_1 *recurre* de nuevo como variable de ramificación, y los dos nuevos subproblemas creados tienen $x_1 = 0$ (debido a la nueva restricción $x_1 \leq 0$ y la restricción de no negatividad sobre x_1) y $x_1 = 1$ (debido a la nueva restricción $x_1 \geq 1$ y la anterior $x_1 \leq 1$).

El tercer cambio se hace en el *paso de acotamiento*. Antes, con un problema de PE *pura* y coeficientes enteros en la función objetivo, el valor de Z de la solución óptima del relajamiento de PL del subproblema se *redondeaba hacia abajo* para obtener la cota, puesto que cualquier solución factible del problema debía tener una Z *entera*. Ahora, con algunas variables *sin* la restricción de integridad, la cota es el valor de Z *sin* redondear.

FIGURA 11.10
Ilustración del fenómeno de una *variable de ramificación recurrente*, donde x_1 se convierte en una variable de ramificación tres veces por tener un valor no entero en la solución óptima del relajamiento de PL en tres nodos.



Recuadro de aplicación

Waste Management, **Inc.** (una de las 100 empresas de Fortune), con casa matriz en Houston, Texas, es un proveedor líder de servicios de manejo de basura en Norteamérica. Su red de operaciones está formada por 293 sitios para depósito de basura, 16 plantas de conversión de basura en energía, 72 instalaciones para la conversión de gas en energía, 146 plantas de reciclaje, 346 estaciones de transferencia y 435 plantas de recolección (depósitos) para brindar servicios a casi 200 millones de clientes residenciales y 2 millones de clientes comerciales en Estados Unidos y Canadá.

Es necesario que los vehículos de recolección y transferencia de la compañía recorran casi 20 000 rutas todos los días. Con un costo de operación anual de casi 120 000 dólares por vehículo, la gerencia deseaba contar con un sistema de administración de rutas que fuera capaz de maximizar la eficiencia de cada una de las rutas. Por tanto, se formó un grupo de IO con el fin de atacar este problema.

La parte medular del sistema de administración de rutas que desarrolló dicho grupo es un enorme *modelo BIP mezclado* que optimiza las rutas asignadas a los respectivos vehículos de recolección y transferencia. Como la función objetivo toma en cuenta varios factores, el objetivo principal consiste en minimizar el tiempo total de viaje. Las variables de decisión principales son variables binarias iguales a 1 si la ruta asignada a un vehículo particular incluye una posible variación e iguales a 0 de otra forma. Un sistema de información geográfica (GIS) proporciona información acerca de la distancia y el tiempo que se requieren para recorrer cualquier par de puntos. Todo esto se encuentra incluido en la aplicación Java basada en la Web que está integrada en los demás sistemas de la compañía.

Se estima que la impantación reciente de este sistema de administración de rutas *incrementará el flujo de efectivo de la compañía* en **648 millones de dólares** en un periodo de 5 años, en gran medida debido a los ahorros del orden de **498 millones de dólares** en gastos de operación en el mismo periodo. Asimismo, brinda un mejor servicio a los clientes.

Fuente: S. Sahoo, S. Kim, B. I. Kim, B. Krass y A. Popov, Jr.: "Routing Optimization for Waste Management", en *Interfaces*, **35**(1): 24-36, enero-febrero de 2005. (En el sitio en internet de este libro —www.mhhe.com/hillier— se puede encontrar una liga hacia este artículo.)

El cuarto (y último) cambio al algoritmo de PEB para obtener el algoritmo de PEM se refiere a la prueba de sondeo 3. Antes, con un problema de PE *pura*, la prueba consistía en que la solución al relajamiento de PL del subproblema fuera *entera*, puesto que ello aseguraba una solución factible y, por tanto, óptima del subproblema. Ahora, con un problema de PE *mixta*, la prueba requiere que sólo las variables *restringidas a enteras* tengan valores *enteros* en la solución óptima del relajamiento de PL del subproblema, puesto que esto es suficiente para asegurar que la solución es factible y, por tanto, óptima para el subproblema.

La incorporación de estos cuatro cambios al resumen que se presentó en la sección anterior sobre el algoritmo de programación entera binaria conduce al siguiente resumen del nuevo algoritmo de programación entera mixta. (Como antes, este resumen supone que se va a maximizar la función objetivo, pero el único cambio necesario para la minimización consiste en cambiar la dirección de la desigualdad de la prueba de sondeo 1.)

Resumen del algoritmo de ramificación y acotamiento de PEM

Paso inicial: Se establece $Z^* = -\infty$. Se aplica el paso de acotamiento, el paso de sondeo y la prueba de optimalidad que se describe después del problema completo. Si no queda sondeado, se clasifica este problema como el único subproblema restante para realizar la primera iteración completa.

Pasos de cada iteración:

1. *Ramificación:* Entre los subproblemas *restantes* (no sondeados), se selecciona el de *creación más reciente*. (Los empates se rompen con la *cota más grande*.) Entre las variables *restringidas a enteros*, que tienen valores *no enteros* en la solución óptima del relajamiento de PL del subproblema, se elige *la primera* en el orden natural como la *variable de ramificación*. Si x_j es esta variable y x_j^* su valor en esta solución, se debe ramificar desde el nodo del subproblema para crear dos nuevos subproblemas luego de agregar las restricciones respectivas $x_j \leq [x_j^*]$ y $x_j \geq [x_j^*] + 1$.
2. *Acotamiento:* Se obtiene la cota de cada subproblema si se aplica el método símplex (o el método símplex dual si se reoptimiza) al relajamiento de PL y se utiliza el valor de Z para la solución óptima resultante.

3. *Sondeo:* Se aplican las pruebas de sondeo que se presentan a continuación a cada nuevo subproblema y se descartan aquellos que quedan sondeados por cualquiera de las pruebas.
Prueba 1: Su cota $\leq Z^*$, donde Z^* es el valor de Z en la solución *de apoyo* actual.
Prueba 2: Su relajamiento de PL no tiene soluciones factibles.
Prueba 3: La solución óptima para su relajamiento de PL tiene valores *enteros* en todas sus variables *restringidas a enteros*. (Si esta solución es mejor que la de apoyo, se convierte en la nueva solución de apoyo y se vuelve a aplicar la prueba 1 con la nueva Z^* a todos los subproblemas no sondeados.)

Prueba de optimalidad: El proceso se detiene cuando *no* hay subproblemas restantes; la solución *de apoyo actual* es óptima.[8] De otra manera, se realiza otra iteración.

Ejemplo de PEM. Ahora se ilustrará este algoritmo mediante su aplicación al siguiente problema de PEM:

$$\text{Maximizar} \quad Z = 4x_1 - 2x_2 + 7x_3 - x_4,$$

sujeta a

$$\begin{aligned} x_1 \quad\quad + 5x_3 \quad\quad &\leq 10 \\ x_1 + x_2 - x_3 \quad\quad &\leq 1 \\ 6x_1 - 5x_2 \quad\quad\quad\quad &\leq 0 \\ -x_1 \quad\quad + 2x_3 - 2x_4 &\leq 3 \end{aligned}$$

y

$$x_j \geq 0, \quad \text{para } j = 1, 2, 3, 4$$
$$x_j \text{ es entero}, \quad \text{para } j = 1, 2, 3.$$

Observe que el número de variables restringidas a enteros es $I = 3$, de manera que x_4 es la única variable continua.

Paso inicial. Después de establecer $Z^* = -\infty$, se forma el relajamiento de PL de este problema mediante la *eliminación* del conjunto de restricciones si x_j es entero para $j = 1, 2, 3$. Si se aplica el método símplex a este relajamiento de PL, la solución óptima es

$$\text{Relajamiento de PL de todo el problema: } (x_1, x_2, x_3, x_4) = \left(\frac{5}{4}, \frac{3}{2}, \frac{7}{4}, 0\right), \quad \text{con } Z = 14\frac{1}{4}.$$

Como tiene soluciones *factibles* y esta solución óptima tiene valores *no enteros* para sus variables restringidas a enteros, se sondea todo el problema y el algoritmo continúa abajo con la primera iteración completa.

Iteración 1. En esta solución óptima del relajamiento de PL la *primera* variable restringida a enteros que tiene un valor no entero es $x_1 = \frac{5}{4}$, con lo cual ésta se convierte en la variable de ramificación. Cuando se ramifica desde el nodo de *todo* el problema (soluciones factibles de *todo*) con esta variable se crean los siguientes dos subproblemas:

Subproblema 1:
Problema original más la restricción adicional

$$x_1 \leq 1.$$

[8] Si no hay solución de apoyo (incumbente), la conclusión es que el problema no tiene soluciones factibles.

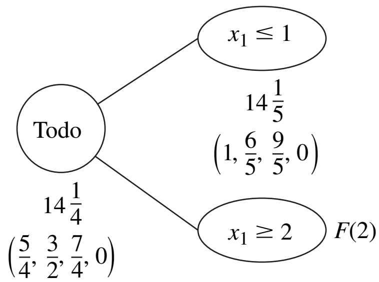


FIGURA 11.11
Árbol de solución después de la primera iteración del algoritmo de ramificación y acotamiento de PEM del ejemplo.

Subproblema 2:
Problema original más la restricción adicional

$$x_1 \geq 2.$$

De nuevo se elimina el conjunto de restricciones a valores enteros y se resuelven los relajamientos de PL de estos dos subproblemas; los resultados son:

Subproblema 1:

Solución óptima para la relajación PL: $(x_1, x_2, x_3, x_4) = \left(1, \frac{6}{5}, \frac{9}{5}, 0\right)$, con $Z = 14\frac{1}{5}$.

Cota: $Z \leq 14\frac{1}{5}$.

Subproblema 2:
Relajación PL: Soluciones no factibles.

Este resultado del subproblema 2 significa que queda sondeado por la prueba 2. Igual que en el caso del problema completo, el subproblema 1 no pasa las pruebas de sondeo.

En la figura 11.11 se resumen todos estos resultados en el árbol de solución.

Iteración 2. Con sólo un subproblema restante que corresponde al nodo $x_1 \leq 1$ en la figura 11.11, la siguiente ramificación se hace desde ahí. Cuando se examina la solución óptima del relajamiento de PL que se presenta en seguida, se observa que este nodo revela que la *variable de ramificación* es x_2, puesto que $x_2 = \frac{6}{5}$ es la primera variable restringida a enteros que no tiene un valor entero. Al agregar una de las restricciones, $x_2 \leq 1$ o $x_2 \geq 2$, se crean los dos nuevos subproblemas que siguen.

Subproblema 3:
Problema original más las restricciones adicionales

$$x_1 \leq 1, \quad x_2 \leq 1.$$

Subproblema 4:
Problema original más las restricciones adicionales

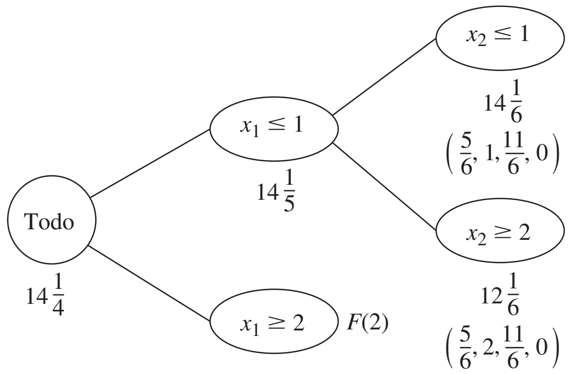
$$x_1 \leq 1, \quad x_2 \geq 2.$$

Al resolver sus relajamientos de PL se obtienen los siguientes resultados.

Subproblema 3:

Solución óptima de la relajación PL: $(x_1, x_2, x_3, x_4) = \left(\frac{5}{6}, 1, \frac{11}{6}, 0\right)$, con $Z = 14\frac{1}{6}$.

Cota: $Z \leq 14\frac{1}{6}$.



■ **FIGURA 11.12**
Árbol de solución después de la segunda iteración del algoritmo de ramificación y acotamiento de PEM del ejemplo.

Subproblema 4:

Solución óptima de la relajación LP: $(x_1, x_2, x_3, x_4) = \left(\frac{5}{6}, 2, \frac{11}{6}, 0\right)$, con $Z = 12\frac{1}{6}$.

Cota: $Z \leq 12\frac{1}{6}$.

Como ambas soluciones existen (soluciones factibles) y tienen valores no enteros de variables restringidas a enteros, ninguno de los subproblemas se sondea. (La prueba 1 no es operativa, puesto que todavía $Z^* = -\infty$, hasta que se encuentre la primera solución de apoyo.)

Hasta este punto, el árbol de solución se presenta en la figura 11.12.

Iteración 3. De los dos subproblemas restantes (3 y 4) que se crearon al mismo tiempo, se selecciona el que tiene la cota más grande (subproblema 3, con $14\frac{1}{6} > 12\frac{1}{6}$) para la siguiente ramificación. Como $x_1 = \frac{5}{6}$ tiene un valor no entero en la solución óptima del relajamiento de PL de su subproblema, x_1 se convierte en la variable de ramificación. (Observe que x_1 es ahora una variable de ramificación *recurrente*, pues también se seleccionó en la iteración 1.) Esto conduce a los siguientes subproblemas.

Subproblema 5:
Problema original más las restricciones adicionales

$x_1 \leq 1$
$x_2 \leq 1$
$x_1 \leq 0$ (entonces $x_1 = 0$).

Subproblema 6:
Problema original más las restricciones adicionales

$x_1 \leq 1$
$x_2 \leq 1$
$x_1 \geq 1$ (entonces $x_1 = 1$).

Los resultados que se obtienen cuando se resuelven los relajamientos de PL son los siguientes.

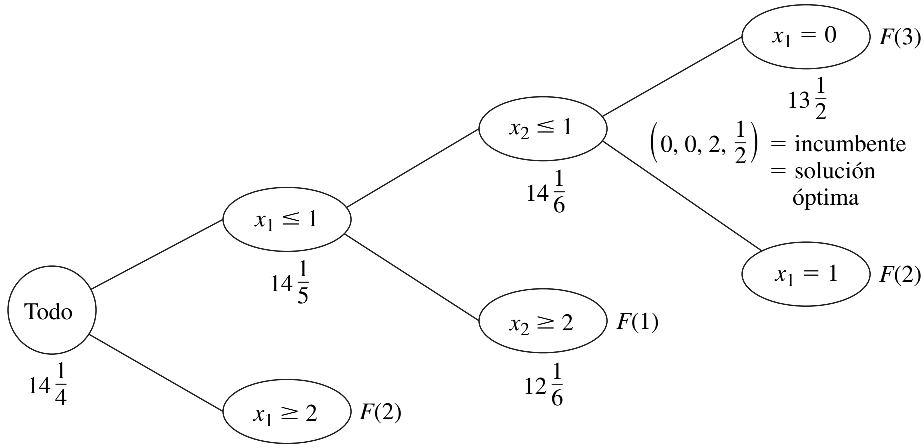
Subproblema 5:

Solución óptima de la relajación LP: $(x_1, x_2, x_3, x_4) = \left(0, 0, 2, \frac{1}{2}\right)$, con $Z = 13\frac{1}{2}$.

Cota: $Z \leq 13\frac{1}{2}$.

Subproblema 6:

Relajamiento LP No tiene soluciones factibles.



■ **FIGURA 11.13**
Árbol de solución después de la última (tercera) iteración del algoritmo de ramificación y acotamiento de PEM del ejemplo.

El subproblema 6 se sondea de inmediato con la prueba 2. Sin embargo, observe que el subproblema 5 también se puede sondear. La prueba 3 pasa porque la solución óptima de este relajamiento de PL tiene valores enteros ($x_1 = 0$, $x_2 = 0$, $x_3 = 2$) para las tres variables restringidas a enteros. (No importa que $x_4 = \frac{1}{2}$, puesto que x_4 no está restringida a enteros.) Esta solución *factible* para el problema original se convierte en la primera solución de apoyo (incumbente):

$$\text{Incumbente} = \left(0,\ 0,\ 2,\ \frac{1}{2}\right) \qquad \text{con } Z^* = 13\frac{1}{2}.$$

Con esta Z^* se vuelve a realizar la prueba de sondeo 1 al otro subproblema (subproblema 4) y pasa la prueba, puesto que su cota $12\frac{1}{6} \le Z^*$.

Esta iteración tuvo éxito en el sondeo de las tres maneras posibles. Lo que es más, ya no hay subproblemas restantes; por tanto, la solución de apoyo actual es óptima.

$$\text{Solución óptima} = \left(0,\ 0,\ 2,\ \frac{1}{2}\right), \quad \text{con } Z = 13\frac{1}{2}.$$

Estos resultados se resumen en el árbol de solución de la figura 11.13.

En el OR Tutor se presenta **otro ejemplo** de aplicación del algoritmo de PEM. Además, en la sección Worked Examples en el sitio en internet de este libro se presenta un **ejemplo pequeño** (sólo dos variables, pero restringidas a enteros) que incluye presentaciones gráficas. En el IOR Tutorial se incluye también una rutina interactiva para ejecutar este algoritmo.

11.8 ENFOQUE DE RAMIFICACIÓN Y CORTE PARA RESOLVER PROBLEMAS DE PEB

La programación entera ha sido un área emocionante de IO durante los últimos años debido a los notables adelantos que se han logrado en su metodología de solución.

Antecedentes

Para dar una mejor perspectiva sobre este progreso, considere los antecedentes históricos. A finales de la década de 1960 y principios de la de 1970 se presentó un gran cambio con el desarrollo y perfeccionamiento del enfoque de ramificación y acotamiento. Después, todo siguió igual; se podían resolver en forma muy eficiente problemas relativamente pequeños (muy por debajo de las 100 variables) pero aún un pequeño aumento del tamaño del problema podía causar una explosión en el tiempo de cálculos más allá de los límites factibles. Se hacían pocos progresos para vencer

este crecimiento exponencial del tiempo de cálculos cuando el tamaño del problema aumentaba. Muchos problemas importantes que surgían en la práctica no se podían resolver.

Más tarde llegó el siguiente cambio, a mediados de la década de 1980, con la introducción del *enfoque de ramificación y corte* para resolver problemas de PEB. Hubo reportes iniciales de problemas muy grandes con hasta 2 000 variables que se pudieron resolver mediante el uso de esta herramienta. Este avance creó un gran interés y condujo a una extensa investigación y actividades de desarrollo para perfeccionar el enfoque que se había utilizado desde entonces. Al principio, el enfoque se limitó a PEB *pura*, pero después se extendió a PEB *mixta*, y con posterioridad a problemas de PEM con algunas variables enteras generales. La descripción que se presentará aquí se limita al enfoque del caso de la PEB *pura*.

En la actualidad es bastante común utilizar el enfoque de ramificación y corte para resolver problemas con unos cuantos miles de variables, y en ocasiones hasta cientos de miles de variables. Como se mencionó en la sección 11.4, esta impresionante capacidad se debe al enorme progreso en tres áreas —mejoras sorprendentes en los algoritmos BIP luego de la incorporación y desarrollo del método de ramificar y cortar, mejoras significativas en los algoritmos de programación lineal que se utilizan ampliamente dentro de los algoritmos BIP y la gran velocidad a la que operan las computadoras (entre ellas, las de escritorio).

Es necesario agregar una nota de precaución. Este enfoque algorítmico no puede resolver siempre *todos* los problemas de PEB pura con unos cuantos miles, o cientos, de variables. Los problemas muy grandes de PEB pura que se ha resuelto tenían matrices **A** *ralas*, es decir, el porcentaje de coeficientes en las restricciones funcionales *distintos de cero* era bastante pequeño (quizá menos de 5%, o incluso menos de 1%). En realidad, en gran medida el enfoque depende de lo ralo. (Por fortuna, las formas ralas son frecuentes en los problemas prácticos.) Más aún, existen otros factores importantes, además de lo ralo y el tamaño, que incrementan o reducen la dificultad de la solución de ciertos problemas de programación entera. Por ello, debe tenerse mucho cuidado al desarrollar las formulaciones de PE de tamaño bastante grande.

Aunque está más allá del alcance de este libro describir con detalle el nuevo enfoque algorítmico, se dará una idea global breve. Como esta idea global se limita a la PEB *pura*, *todas* las variables que se introducen en adelante en esta sección son variables *binarias*.

Este enfoque utiliza en forma principal una combinación de los tres tipos de técnicas:[9] el *preprocesado automático del problema*, la *generación de planos cortantes* o *cortaduras* y las ingeniosas técnicas de *ramificación y acotamiento*. Usted está familiarizado con las técnicas de ramificación y acotamiento y no se profundizará más en las versiones más elaboradas que se incorporan aquí. A continuación se presenta una introducción conceptual sobre los otros dos tipos de técnicas que se utilizan.

Preprocesado automático del problema de PEB pura

El preprocesado automático del problema incluye una “inspección por computadora” de la formulación del problema de PE que proporciona el usuario, con el fin de detectar algunas reformulaciones que permiten resolver el problema con más rapidez, sin eliminar ninguna solución factible. Estas reformulaciones se pueden clasificar en tres categorías:

1. *Fijación de variables:* Identificar las variables que pueden fijarse en uno de sus dos valores posibles (0 o 1) debido a que el otro valor no puede ser parte de una solución que sea tanto factible como óptima.
2. *Eliminación de restricciones redundantes:* Identificar y eliminar *restricciones redundantes* (restricciones que son satisfechas en forma automática cuando las soluciones satisfacen todas las demás).
3. *Estrechamiento de restricciones:* Estrechar algunas restricciones para reducir la región factible del relajamiento de PL sin eliminar ninguna solución factible del problema de PEB.

Estas categorías se describirán más adelante.

[9] Como se explicó de manera breve en la sección 11.4, existe una técnica más que ha jugado un papel significativo en el progreso reciente, que es el uso de la *heurística* para encontrar soluciones factibles buenas con rapidez.

Fijación de variables. El principio general para fijar variables es el siguiente.

> Si el valor de una variable no puede satisfacer cierta restricción, aun cuando las otras variables sean iguales a sus mejores valores que tratan de satisfacer la restricción, entonces esa variable debe fijarse en el otro valor.

Por ejemplo, *cada* una de las siguientes restricciones de tipo ≤ permitiría fijar $x_1 = 0$, puesto que $x_1 = 1$ con los mejores valores de las otras variables (0 con un coeficiente no negativo y 1 con un coeficiente negativo) violaría la restricción.

$$
\begin{aligned}
3x_1 \le 2 \quad &\Rightarrow \quad x_1 = 0, \quad \text{puesto que} \quad 3(1) > 2.\\
3x_1 + x_2 \le 2 \quad &\Rightarrow \quad x_1 = 0, \quad \text{puesto que} \quad 3(1) + 1(0) > 2.\\
5x_1 + x_2 - 2x_3 \le 2 \quad &\Rightarrow \quad x_1 = 0, \quad \text{puesto que} \quad 5(1) + 1(0) - 2(1) > 2.
\end{aligned}
$$

El procedimiento general para verificar cualquier restricción de tipo ≤ es identificar la variable con el *coeficiente positivo mayor*, y si la *suma de ese coeficiente* y cualesquiera *coeficientes negativos* excede la cantidad del lado derecho, entonces esa variable debe fijarse en 0. (Una vez que se fija la variable el procedimiento se puede repetir para la variable con el siguiente coeficiente positivo más grande, y así sucesivamente.)

Un procedimiento análogo con restricciones de tipo ≥ permite fijar una variable en 1 como se ilustra tres veces en seguida.

$$
\begin{aligned}
3x_1 \ge 2 \quad &\Rightarrow \quad x_1 = 1, \quad \text{puesto que} \quad 3(0) < 2.\\
3x_1 + x_2 \ge 2 \quad &\Rightarrow \quad x_1 = 1, \quad \text{puesto que} \quad 3(0) + 1(1) < 2.\\
3x_1 + x_2 - 2x_3 \ge 2 \quad &\Rightarrow \quad x_1 = 1, \quad \text{puesto que} \quad 3(0) + 1(1) - 2(0) < 2.
\end{aligned}
$$

Una restricción de tipo ≥ también puede permitir que se fije una variable en 0, como se ilustra a continuación.

$$x_1 + x_2 - 2x_3 \ge 1 \quad \Rightarrow \quad x_3 = 0, \quad \text{puesto que} \quad 1(1) + 1(1) - 2(1) < 1.$$

El siguiente ejemplo muestra una restricción de tipo ≥ en la que se fija una variable en 1 y otra en 0.

$$
\begin{aligned}
3x_1 + x_2 - 3x_3 \ge 2 \quad &\Rightarrow \quad x_1 = 1, \quad \text{puesto que} \quad 3(0) + 1(1) - 3(0) < 2\\
y \quad &\Rightarrow \quad x_3 = 0, \quad \text{puesto que} \quad 3(1) + 1(1) - 3(1) < 2.
\end{aligned}
$$

De igual manera, una restricción de tipo ≤ con un valor *negativo* en el lado derecho puede dar como resultado que 0 o 1 se conviertan en el valor fijo de una variable. Por ejemplo, ocurren ambas cosas con la siguiente restricción:

$$
\begin{aligned}
3x_1 - 2x_2 \le -1 \quad &\Rightarrow \quad x_1 = 0, \quad \text{puesto que} \quad 3(1) - 2(1) > -1\\
y \quad &\Rightarrow \quad x_2 = 1, \quad \text{puesto que} \quad 3(0) - 2(0) > -1.
\end{aligned}
$$

El hecho de fijar una variable a partir de una restricción puede llegar a generar una reacción en cadena que permita después fijar otras variables a partir de otras restricciones. Por ejemplo, vea lo que ocurre con las tres restricciones siguientes.

$$3x_1 + x_2 - 2x_3 \ge 2 \quad \Rightarrow \quad x_1 = 1 \quad \text{(como antes)}.$$

Entonces

$$x_1 + x_4 + x_5 \le 1 \quad \Rightarrow \quad x_4 = 0, \quad x_5 = 0.$$

En consecuencia,

$$-x_5 + x_6 \le 0 \quad \Rightarrow \quad x_6 = 0.$$

En algunos casos es posible combinar una o más restricciones de *alternativas mutuamente excluyentes* con otra restricción para fijar una variable, como se verá en seguida:

$$
\left.\begin{aligned}
8x_1 - 4x_2 - 5x_3 + 3x_4 &\le 2\\
x_2 + x_3 \phantom{{}+ 3x_4} &\le 1
\end{aligned}\right\} \quad \Rightarrow \quad x_1 = 0,
$$

$$\text{puesto que} \quad 8(1) - \text{máx}\{4, 5\}(1) + 3(0) > 2.$$

Existen otras técnicas para fijar variables, incluso algunas que toman en cuenta ciertos factores de optimalidad, pero no se profundizará más en este tema.

La fijación de variables puede tener un efecto muy marcado en la reducción del tamaño de un problema. No es inusual eliminar hasta la mitad de las variables del problema para su consideración posterior.

Eliminación de restricciones redundantes. A continuación se presenta una manera sencilla de detectar una restricción redundante.

> Si una restricción funcional satisface incluso la solución binaria de mayor reto, entonces las restricciones binarias las han convertido en redundantes y en adelante pueden eliminarse. En el caso de una restricción de tipo $\leq$ la solución binaria más demandante tiene variables iguales a 1 cuando tiene coeficientes no negativos y otras variables son iguales a 0. (Estos valores se invierten si se trata de una restricción de tipo $\geq$.)

A continuación se presentan algunos ejemplos.

$3x_1 + 2x_2 \leq 6$ es redundante porque $3(1) + 2(1) \leq 6$.
$3x_1 - 2x_2 \leq 3$ es redundante porque $3(1) - 2(0) \leq 3$.
$3x_1 - 2x_2 \geq -3$ es redundante porque $3(0) - 2(1) \geq -3$.

En la mayoría de los casos, cuando se identifica una restricción como redundante, en el modelo original no era redundante sino que se convirtió después de fijar algunas variables. De los 11 ejemplos que se dieron para fijar variables, *todos* menos el último dejaron una restricción que en ese momento se convirtió en redundante.

Estrechamiento de restricciones.[10] Considere el siguiente problema.

Maximizar $Z = 3x_1 + 2x_2$,

sujeta a

$$2x_1 + 3x_2 \leq 4$$

y

x_1, x_2 binaria.

Este problema de PEB tiene sólo tres soluciones factibles —(0, 0), (1, 0) y (0, 1)— de las cuales la solución óptima es (1, 0) con $Z = 3$. La región factible del relajamiento de PL de este problema se muestra en la figura 11.14. La solución óptima de este relajamiento de PL es $(1, \frac{2}{3})$ con $Z = 4\frac{1}{3}$, que no está muy cerca de la solución óptima del problema de PEB. Un algoritmo de ramificación y acotamiento tendrá que trabajar bastante para identificar la solución óptima.

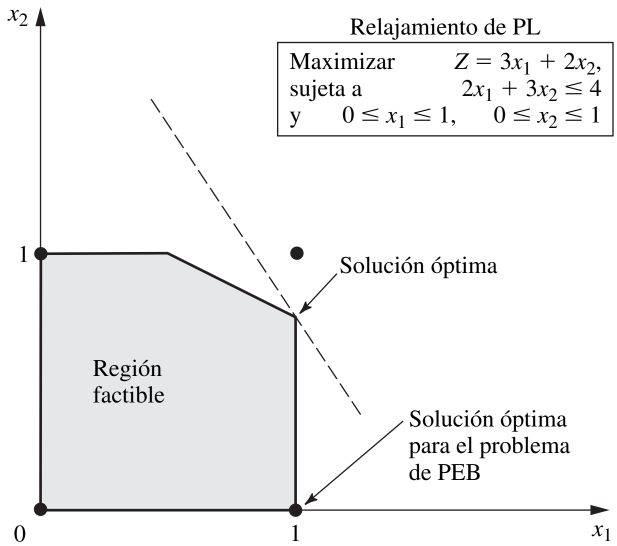
Ahora observe lo que ocurre si la restricción funcional $2x_1 + 3x_2 \leq 4$ se sustituye por

$$x_1 + x_2 \leq 1.$$

Las soluciones factibles para el problema de PEB son exactamente las mismas —(0, 0), (1, 0) y (0, 1)— de manera que la solución óptima todavía es (1, 0). Sin embargo, la región factible para el relajamiento de PL se ha reducido de manera considerable, como se muestra en la figura 11.15. De hecho, esta región factible se ha reducido tanto que la solución óptima del relajamiento de PL es ahora (1, 0), de manera que se ha encontrado la solución óptima del problema de PEB sin necesidad de trabajo adicional.

Éste es un ejemplo de cómo se estrecha una restricción de manera que se reduzca la región factible del relajamiento de PL sin eliminar ninguna de las soluciones factibles del problema de PEB. Fue sencillo realizarlo para este problema pequeño de dos variables que se podía dibujar en una gráfica. No obstante, con la aplicación de los mismos principios para estrechar restricciones sin eliminar ninguna de las soluciones factibles de PEB, se puede usar el siguiente procedimiento algebraico para cualquier restricción de tipo $\leq$ con cualquier número de variables.

[10] También se conoce comúnmente como *reducción de coeficientes*.



■ **FIGURA 11.14**
Relajamiento de PL (incluye su región factible y la solución óptima) del ejemplo de PEB usado para ilustrar cómo estrechar una restricción.

Procedimiento para estrechar o reducir una restricción de tipo $\leq$

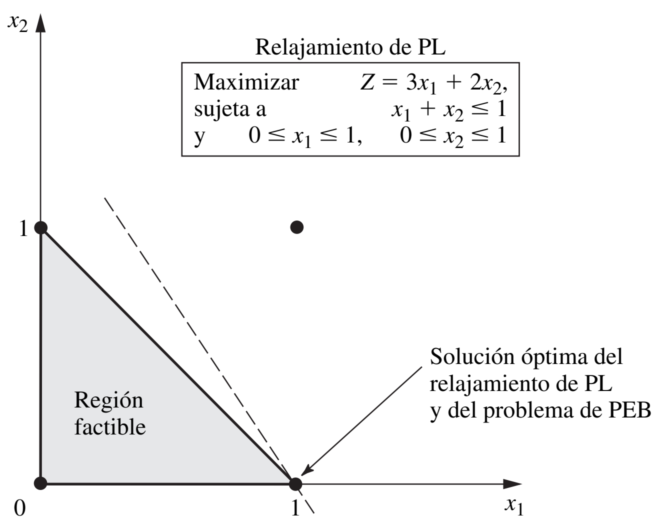
Denote la restricción por $a_1x_1 + a_2x_2 + \cdots + a_nx_n \leq b$.

1. Calcule S = suma de los a_j *positivos*.
2. Identifique cualquier $a_j \neq 0$ tal que $S < b + |a_j|$.
   a) Si no existen, el procedimiento se detiene; la restricción ya no se puede reducir.
   b) Si $a_j > 0$, vaya al paso 3.
   c) Si $a_j < 0$, vaya al paso 4.
3. ($a_j > 0$) Calcule $\bar{a}_j = S - b$ y $\bar{b} = S - a_j$. Restablezca $a_j = \bar{a}_j$ y $b = \bar{b}$. Regrese al paso 1.
4. ($a_j < 0$) Aumente a_j a $a_j = b - S$. Regrese al paso 1.

La aplicación de este procedimiento a la restricción funcional del ejemplo anterior fluye de la siguiente manera:

La restricción es $2x_1 + 3x_2 \leq 4$ $\quad (a_1 = 2, a_2 = 3, b = 4)$.

■ **FIGURA 11.15**
Relajamiento de PL después de estrechar la restricción $2x_1 + 3x_2 \leq 4$, a $x_1 + x_2 \leq 1$ del ejemplo de la figura 11.14.



1. $S = 2 + 3 = 5$.
2. a_1 satisface $S < b + |a_1|$, puesto que $5 < 4 + 2$. Además a_2 satisface $S < b + |a_2|$, puesto que $5 < 4 + 3$. Seleccione a_1 en forma arbitraria.
3. $\bar{a}_1 = 5 - 4 = 1$ y $\bar{b} = 5 - 2 = 3$, se restablece $a_1 = 1$ y $b = 3$. La nueva restricción reducida es

$$x_1 + 3x_2 \le 3 \qquad (a_1 = 1, a_2 = 3, b = 3).$$

1. $S = 1 + 3 = 4$.
2. a_2 satisface $S < b + |a_2|$, puesto que $4 < 3 + 3$.
3. $\bar{a}_2 = 4 - 3 = 1$ y $\bar{b} = 4 - 3 = 1$, se restablece $a_2 = 1$ y $b = 1$. La nueva restricción reducida o estrecha es

$$x_1 + x_2 \le 1 \qquad (a_1 = 1, a_2 = 1, b = 1).$$

1. $S = 1 + 1 = 2$.
2. Ninguna $a_j \neq 0$ satisface $S < b + |a_j|$, por lo cual el procedimiento se detiene; $x_1 + x_2 \le 1$ es la restricción reducida que se busca.

Si al ejecutar el paso 2 la primera vez, en el ejemplo anterior, se hubiera seleccionado a_2, entonces la primera restricción reducida habría sido $2x_1 + x_2 \le 2$. La siguiente serie de pasos de nuevo habrían conducido a $x_1 + x_2 \le 1$.

El siguiente es otro ejemplo en el que el procedimiento estrecha la restricción de la izquierda y la convierte en la de la derecha, y después reduce más a la segunda de la derecha.

$$4x_1 - 3x_2 + x_3 + 2x_4 \le 5 \quad \Rightarrow \quad 2x_1 - 3x_2 + x_3 + 2x_4 \le 3$$
$$\Rightarrow \quad 2x_1 - 2x_2 + x_3 + 2x_4 \le 3.$$

(El problema 11.8-5 pide que se aplique el procedimiento para confirmar estos resultados.)

Una restricción de la forma $\ge$ se puede convertir a la forma $\le$ (al multiplicar ambos lados por -1) para aplicar este procedimiento de manera directa.

Generación de planos cortantes para PEB pura

Un **plano cortante** (o **corte**) de un problema de PE es una nueva restricción funcional que reduce la región factible del relajamiento de PL sin eliminar soluciones factibles del problema de PE original. De hecho, se acaba de estudiar una manera de generar planos cortantes en problemas de PE pura, esto es, la aplicación del procedimiento anterior para estrechar restricciones. De esta forma, $x_1 + x_2 \le 1$ es un plano cortante del problema PEB que se presenta en la figura 11.14, que conduce a la región factible reducida del relajamiento de PL que se muestra en la figura 11.15.

Además de este procedimiento, se ha desarrollado un buen número de técnicas para generar planos cortantes que tenderán a acelerar la rapidez con la que el método de ramificación y acotamiento puede encontrar una solución óptima a un problema de PEB pura. Se centrará la atención nada más en una de estas técnicas.

Para ilustrar esta técnica, considere el problema de PE pura de la California Manufacturing Co., que se presentó en la sección 11.1 y que se utilizó para ilustrar el algoritmo de ramificación y acotamiento de PEB en la sección 11.6. La solución óptima de su relajamiento de PL se presenta en la figura 11.5 como $(x_1, x_2, x_3, x_4) = (\frac{5}{6}, 1, 0, 1)$. Una de las restricciones funcionales es

$$6x_1 + 3x_2 + 5x_3 + 2x_4 \le 10.$$

Ahora observe que, juntas, las restricciones binarias y esta restricción implican que

$$x_1 + x_2 + x_4 \le 2.$$

Esta nueva restricción es un *plano cortante*. Elimina parte de la región factible del relajamiento de PL, incluso la que llegó a ser la solución óptima, $(\frac{5}{6}, 1, 0, 1)$, pero no elimina soluciones *enteras* factibles. Cuando se agrega este corte al modelo original, mejora el desempeño del algoritmo de ramificación y acotamiento de PEB de la sección 11.6 (vea la figura 11.9) en dos formas. Primero, la solución óptima del nuevo (y más estrecho) relajamien-

to de PL sería $(1, 1, \frac{1}{5}, 0)$, con $Z = 15\frac{1}{5}$, de manera que las cotas del nodo de *Todo*, el nodo $x_1 = 1$ y el nodo $(x_1, x_2) = (1, 1)$, serían ahora de 15 en lugar de 16. Segundo, se necesitaría una iteración menos, porque la solución óptima del relajamiento de PL en el nodo $(x_1, x_2, x_3) = (1, 1, 0)$ ahora sería (1, 1, 0, 0), que proporciona una nueva *solución de apoyo* con $Z^* = 14$. Por tanto, en la *tercera* iteración (vea la figura 11.8) se sondearía este nodo por la prueba 3 y el nodo $(x_1, x_2) = (1, 0)$ se sondearía por la prueba 1, lo cual indicaría que esta nueva solución de apoyo es la solución óptima del problema de PEB original.

El siguiente es el procedimiento general que se utiliza para generar este corte.

Procedimiento para generar planos cortantes

1. Considere cualquier restricción funcional de la forma $\leq$ con sólo coeficientes no negativos.
2. Encuentre un grupo de variables (llamado **cubierta mínima** de la restricción) tales que
   a) La restricción se viola si todas las variables del grupo son iguales a 1 y todas las demás variables son iguales a 0.
   b) Pero la restricción se satisface si el valor de *cualquiera* de estas variables cambia de 1 a 0.
3. Si N denota el número de variables del grupo, el plano cortante obtenido tiene la forma

   Suma de variables del grupo $\leq N - 1$.

Al aplicar este procedimiento a la restricción $6x_1 + 3x_2 + 5x_3 + 2x_4 \leq 10$, se ve que el grupo de variables $\{x_1, x_2, x_4\}$ es una *cubierta mínima* porque

a) (1, 1, 0, 1) viola la restricción.
b) Pero la restricción se satisface si el valor de *cualquiera* de estas tres variables cambia de 1 a 0.

Como en este caso $N = 3$, el plano cortante que se obtiene es $x_1 + x_2 + x_4 \leq 2$.

Esta misma restricción tiene también una segunda cubierta mínima $\{x_1, x_3\}$, puesto que (1, 0, 1, 0) viola la restricción, pero tanto (0, 0, 1, 0) como (1, 0, 0, 0) la satisfacen. Por tanto, $x_1 + x_3 \leq 1$ es otro corte válido.

El enfoque de ramificación y cortadura implica la generación de *muchas* cortaduras en forma parecida antes de aplicar las ingeniosas técnicas de ramificación y acotamiento. Los resultados que se obtienen cuando se incluyen cortes pueden ser importantes para estrechar los relajamientos de PL. En algunos casos la *distancia* entre el valor de Z para la solución óptima del relajamiento de PL del problema completo de PEB y el valor de Z para la solución óptima del problema se reduce hasta en 98%.

Irónicamente, los primeros algoritmos que se desarrollaron para programación entera, incluso el celebrado algoritmo que publicó Ralph Gomory en 1958, estaban basados en cortes (generados de otra manera), pero este enfoque resultó ser poco satisfactorio en la práctica (excepto para algunos tipos especiales de problemas). No obstante, estos algoritmos se apoyaban sólo en los cortes. Ahora se sabe que una *combinación* acertada de los cortes y las técnicas de ramificación y acotamiento (junto con el preprocesado automático del problema) proporciona un enfoque algorítmico poderoso para resolver problemas de PEB a gran escala. Ésta es una razón por la que se ha dado el nombre de algoritmo de *ramificación y corte* a este nuevo enfoque.

11.9 INCORPORACIÓN DE LA PROGRAMACIÓN DE RESTRICCIONES

Ninguna presentación de las ideas básicas de programación entera está completa en estos días si no se le introduce un interesante desarrollo reciente —la incorporación de las técnicas de *programación de restricciones*— que promete ampliar en gran medida las capacidades de formular y resolver modelos de programación entera. (Estas mismas técnicas se comienzan a utilizar en áreas relacionadas de programación matemática, en especial optimización combinatoria, pero esta presentación se limitará a su uso central en programación entera.)

Naturaleza de la programación de restricciones

A mediados de la década de 1980, investigadores de la comunidad de ciencias de la computación desarrollaron la programación de restricciones mediante la combinación de ideas de inteligencia

artificial con el desarrollo de lenguajes de programación de computadoras. La meta era contar con un sistema de programación de computadoras flexible que incluyera tanto *variables* como *restricciones* en sus valores, y que al mismo tiempo permitiera la descripción de procedimientos de búsqueda que generara valores factibles de las variables. Cada variable tiene un *dominio* de valores posibles; por ejemplo, $\{2, 4, 6, 8, 10\}$. En lugar de estar limitadas a los tipos de restricciones matemáticas que se utilizan en programación matemática, existe una gran flexibilidad en la manera de establecer las restricciones. En particular, las restricciones pueden ser de alguno de los siguientes tipos.

1. Restricciones matemáticas; por ejemplo, $x + y < z$.
2. Restricciones disyuntivas; por ejemplo, los tiempos de ciertas tareas en el problema que se modela no se pueden sobreponer.
3. Restricciones relacionales; por ejemplo, al menos tres tareas se deben asignar a cierta máquina.
4. Restricciones explícitas; por ejemplo, aunque x y y tengan dominios $\{1, 2, 3, 4, 5\}$, (x, y) debe ser (1, 1), (2, 3) o (4, 5).
5. Restricciones unitarias; por ejemplo, z es un entero entre 5 y 10.
6. Restricciones lógicas; por ejemplo, si x es 5, entonces y está entre 6 y 8.

Cuando se expresan estos tipos de restricciones, la programación de restricciones permite el uso de varias funciones lógicas normales, como IF, AND, OR, NOT, etc. Excel incluye muchas de las mismas funciones lógicas. Ahora, LINGO cuenta con todas las funciones lógicas normales y se puede usar su optimizador global para encontrar una solución óptima global.

Para ilustrar los algoritmos que utiliza la programación de restricciones para generar soluciones factibles, suponga que un problema tiene cuatro variables —x_1, x_2, x_3, x_4— y sus dominios son

$$x_1 \in \{1, 2\}, x_2 \in \{1, 2\}, x_3 \in \{1, 2, 3), x_4 \in \{1, 2, 3, 4, 5\},$$

donde el símbolo $\in$ significa que la variable de la izquierda pertenece al conjunto de la derecha. Suponga también que las restricciones son

(1) Todas estas variables deben tener valores diferentes,
(2) $x_1 + x_3 = 4$.

Por lógica directa, dado que los valores de 1 y 2 deben estar reservados para x_1 y x_2, la primera restricción implica de inmediato que $x_3 \in \{3\}$, lo cual a su vez implica que $x_4 \in \{4, 5\}$. (Este proceso de eliminar valores posibles para variables se conoce como *dominio de reducción*.) Después, como se ha cambiado el dominio de x_3, el proceso de *propagación de la restricción* se aplica a la segunda restricción para implicar que $x_1 \in \{1\}$. De nueva cuenta, esto dispara la primera restricción, de manera que por medio de

$$x_1 \in \{1\}, \quad x_2 \in \{2\}, \quad x_3 \in \{3\}, \quad x_4 \in \{4, 5\}$$

se presenta la lista de las únicas soluciones factibles del problema. Este tipo de *razonamiento de factibilidad* basado en alternar la aplicación de los algoritmos de reducción del dominio y de propagación de la restricción es una parte crucial de la programación de restricciones.

Después de la aplicación de los algoritmos de propagación de la restricción y reducción del dominio a un problema, se utiliza un procedimiento de búsqueda para encontrar soluciones factibles. En el ejemplo anterior, como los dominios de todas las variables se han reducido a un solo valor excepto para x_4, el procedimiento de búsqueda simplemente podría probar los valores $x_4 = 4$ y $x_4 = 5$ para determinar las soluciones factibles completas de ese problema. Sin embargo, por lo general, en el caso de un problema con muchas restricciones y variables, los algoritmos de propagación de la restricción y reducción del dominio no reducen el dominio de cada variable a un solo valor. Por tanto, es necesario escribir un procedimiento de búsqueda que probará diferentes asignaciones de valores para las variables. A medida que se prueban estas asignaciones, se dispara el algoritmo de propagación de la restricción y entonces ocurre la reducción del dominio. El proceso crea un *árbol de búsqueda*, el cual es similar al árbol de ramificación y acotamiento de programación entera.

El proceso global de aplicar la programación de restricciones a complicados problemas de PE (o problemas relacionados) implica el seguimiento de los siguientes tres pasos.

1. Formular un modelo compacto del problema mediante el uso de una variedad de tipos de restricción (la mayoría de los cuales no cumple con el formato de programación entera).
2. Encontrar soluciones factibles de manera eficiente para satisfacer todas las restricciones.
3. Buscar una solución óptima entre estas soluciones factibles.

El poder de la programación de restricciones reside en su gran capacidad para realizar los primeros dos pasos, mientras que la fortaleza principal de la programación entera y sus algoritmos consiste en la realización del tercer paso. De esta forma, idealmente, la programación de restricciones está diseñada para manejar problemas con muchas restricciones que no tienen función objetivo, de forma que la única meta es encontrar una solución factible. Sin embargo, también se puede extender hacia el tercer paso. Un método para hacerlo es *enumerar* las soluciones factibles y calcular el valor de la función objetivo de cada una. No obstante, este enfoque sería muy ineficiente para enfrentar problemas donde existen muchas soluciones factibles. Para contrarrestar esta desventaja, el enfoque común es agregar una restricción que acote en gran medida la función objetivo hasta valores que estén muy cerca de lo que se anticipa como una solución óptima. Por ejemplo, si el objetivo es *maximizar* la función objetivo y se anticipa que el valor Z aproximado de una solución óptima es $Z = 10$, se podría agregar la restricción de que $Z \geq 9$ de manera que las únicas soluciones factibles que restan por ser enumeradas sean aquellas que estén muy cerca de ser óptimas. Después, cada vez que se encuentre una nueva mejor solución durante la búsqueda, la cota sobre Z se puede estrechar aún más para considerar sólo las soluciones factibles que sean al menos tan buenas como la mejor solución actual.

Aunque éste es un enfoque razonable para el tercer paso, uno más atractivo podría ser integrar programación de restricciones y programación entera para que cada una se utilice de manera primordial en los pasos donde es más fuerte: pasos 1 y 2 con programación de restricciones y paso 3 con programación entera. Esta posibilidad es una parte del potencial de la programación de restricciones que se describe a continuación.

Potencial de la programación de restricciones

En la década de 1990, algunas características de la programación de restricciones, entre ellas poderosos algoritmos de solución de restricciones, se incorporaron de manera exitosa a varios lenguajes de programación de propósito general. Este avance condujo a la ciencia computacional cada vez más cerca del Santo Grial de la programación de computadoras; es decir, permitirle al usuario establecer el problema de manera simple para que después la computadora lo resolviera.

Cuando la noticia de este interesante desarrollo comenzó a esparcirse fuera de la comunidad de la ciencia computacional, los estudiosos de la investigación de operaciones comenzaron a darse cuenta del enorme potencial de integrar la programación de restricciones con las técnicas tradicionales de programación entera (así como de otras áreas de programación matemática). La flexibilidad mucho mayor para expresar las restricciones del problema aumentaría en gran medida la capacidad de formular modelos válidos para solucionar problemas complejos. También podría conducir a formulaciones mucho más compactas y directas. Además, al reducir el tamaño de la región factible que debe considerarse cuando se pretende encontrar soluciones factibles de manera eficiente dentro de esta región, los algoritmos de solución de restricciones dentro de la programación de restricciones podrían ayudar a acelerar el progreso de los algoritmos de programación entera para encontrar una solución óptima.

Debido a sus diferencias sustanciales, integrar la programación de restricciones con la programación entera es una tarea muy difícil. En razón de que esta última no reconoce la mayoría de las restricciones de la programación de restricciones, se requiere el desarrollo de procedimientos implantados por computadora para traducir el lenguaje de la programación entera al de programación de restricciones y viceversa. Se han hecho grandes progresos, pero sin duda esta cuestión seguirá vigente como una de las áreas de investigación de IO más activas durante los años por venir.

Para ilustrar la manera en que la programación de restricciones puede simplificar en gran medida la formulación de modelos de programación entera, a continuación se presentarán dos de las “restricciones globales” más importantes de la programación de restricciones. Una **restricción**

global es una restricción que expresa en forma sucinta un patrón global de la relación permisible entre múltiples variables. Por tanto, con frecuencia, una sola restricción global puede reemplazar aquello que requeriría de una gran cantidad de restricciones tradicionales de programación entera, al mismo tiempo que hace que el modelo sea mucho más fácil de leer. Para clarificar la presentación, se usarán ejemplos muy simples que no requieren el uso de programación de restricciones para ilustrar las restricciones globales, pero estos mismos tipos de restricciones también se pueden usar para resolver algunos problemas mucho más complicados.

Restricción todas diferentes

La restricción global *todas diferentes* sólo especifica que todas las variables de un conjunto dado deben tener valores diferentes. Si $x_1, x_2, \ldots, x_n$ son las variables implicadas, la restricción puede escribirse de manera sucinta como

todas diferentes $(x_1, x_2, \ldots, x_n)$

al mismo tiempo que también se especifican los dominios de las variables individuales del modelo. (Estos dominios en forma colectiva necesitan incluir al menos n valores diferentes para fortalecer la restricción de todas diferentes.)

Para ilustrar esta restricción, considere el *problema de asignación* clásico que se presentó en la sección 8.3. Recuerde que este problema implica conceder n asignados a n tareas en una base de uno a uno de manera que se minimice el costo total de estas asignaciones. Aunque el problema de asignación es muy fácil de resolver (como se describió en la sección 8.4), ilustra muy bien la forma en que la restricción de todas diferentes puede simplificar en gran medida la formulación del modelo.

Con la formulación tradicional que se presentó en la sección 8.3, las variables de decisión son las variables binarias

$$x_{ij} = \begin{cases} 1, & \text{si el asignado } i \text{ realiza la tarea } j \\ 0, & \text{si no es así} \end{cases}$$

para $i, j = 1, 2, \ldots, n$. Si por ahora se hace caso omiso de la función objetivo, las restricciones funcionales son las siguientes.

Cada asignado i se asignará a sólo *una* tarea:

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad \text{para } i = 1, 2, \ldots, n.$$

Cada tarea j será realizada por exactamente *un* asignado:

$$\sum_{i=1}^{n} x_{ij} = 1 \qquad \text{para } j = 1, 2, \ldots, n.$$

Así, hay n^2 variables y $2n$ restricciones funcionales.

Ahora se verá el modelo mucho más pequeño que puede proporcionar la programación de restricciones. En este caso, las variables son

y_i = tareas que se encargan al asignado i

para $i = 1, 2, \ldots, n$. Existen n tareas y se enumeran 1, 2, ..., n, de forma que cada una de las y_i variables tiene el dominio $\{1, 2, \ldots, n\}$. Como todos los asignados deben dirigirse a tareas diferentes, esta restricción sobre las variables es descrita en forma precisa por una sola restricción global,

todas diferentes $(y_1, y_2, \ldots, y_n)$.

Por tanto, en lugar de n^2 variables y $2n$ restricciones funcionales, este modelo de programación de restricciones completo (que excluye la función objetivo) tiene sólo n variables y *una sola* restricción (más un dominio para todas las variables).

Ahora se verá cómo la siguiente restricción global también permite incorporar la función objetivo a este pequeño modelo.

Restricción de elemento

La restricción global de *elemento* se utiliza casi siempre para buscar un costo o ganancia asociados con una variable entera. En particular, suponga que una variable y tiene un dominio $\{1, 2, \ldots, n\}$ y que el costo asociado con cada una de estas variables es $c_1, c_2, \ldots, c_n$, respectivamente. En consecuencia, la restricción

$$\text{elemento } (y, [c_1, c_2, \ldots, c_n], z)$$

restringe la variable z a un valor igual a la y-ésima constante de la lista $[c_1, c_2, \ldots, c_n]$. En otras palabras $z = c_y$. Ahora, esta variable z se puede incluir en la función objetivo para proporcionar el costo asociado con y.

Para ilustrar el uso de la restrición de elemento, considere de nuevo el problema de asignación; sea

$$c_{ij} = \text{costo de asignar a } i \text{ la tarea } j$$

para $i, j, = 1, 2, \ldots, n$. El modelo completo de programación de restricciones (que incluye la función objetivo) de este problema es

$$\text{Minimizar } Z = \sum_{i=1}^{n} z_i,$$

sujeta a

$$\text{elemento } (y_i, [c_{i1}, c_{i2}, \ldots, c_{in}], z_i) \qquad \text{para } i = 1, 2, \ldots, n,$$
$$\text{todas diferentes } (y_1, y_2, \ldots y_n),$$
$$y_i \in \{1, 2, \ldots, n\} \qquad \text{para } i = 1, 2, \ldots, n.$$

Este modelo completo ahora tiene $2n$ variables y $(n + 1)$ restricciones (además de un dominio para todas las variables), el cual es mucho más pequeño que el de la formulación de programación entera que se presentó en la sección 8.3. Por ejemplo, cuando $n = 100$, este modelo tiene 200 variables y 101 restricciones, mientras que el modelo tradicional de programación entera tiene 10 000 variables y 200 restricciones funcionales.

Como un ejemplo adicional, reconsidere el ejemplo 2 (violación de la proporcionalidad) que se presentó en la sección 11.4. En este caso, las variables de decisión originales son

$$x_j = \text{número de anuncios de televisión asignados al producto } j$$

para $j = 1, 2, 3$, donde un total de cinco anuncios de televisión se va a asignar a los tres productos. Sin embargo, como las ganancias que se presentan en la tabla 11.3 para valores diferentes de cada x_j no son proporcionales a x_j. En la sección 11.4 se formularon dos modelos alternativos de programación entera con variables binarias auxiliares para manejar este problema. Ambos modelos son bastante complicados.

Un modelo de programación de restricciones que utiliza la restricción de elemento es mucho más directo. Por ejemplo, la ganancia del producto 1 que se muestra en la tabla 11.3 es 0, 1, 3 y 3 para los valores respectivos $x_1 = 0, 1, 2$ y 3. Por tanto, esta ganancia simplemente es z_1 cuando el valor de z_1 está dado por la restricción

$$\text{elemento } (x_1 + 1, [0, 1, 3, 3], z_1).$$

(El primer componente es $x_1 + 1$ en lugar de x_1 porque $x_1 + 1 = 1, 2, 3$ o 4, y es el valor de este componente el que indica la elección de la posición 1, 2, 3 o 4 en la lista [0, 1, 3, 3]). Al proceder de la misma forma con los otros dos productos, el modelo completo es

$$\text{Maximizar } Z = z_1 + z_2 + z_3,$$

sujeta a

$$\text{elemento } (x_1 + 1, [0, 1, 3, 3], z_1),$$
$$\text{elemento } (x_2 + 1, [0, 0, 2, 3], z_2),$$
$$\text{elemento } (x_3 + 1, [0, -1, 2, 4], z_3),$$

$$x_1 + x_2 + x_3 = 5,$$
$$x_j \in \{0, 1, 2, 3\} \qquad \text{para } j = 1, 2, 3.$$

Ahora compare este modelo con los dos modelos de programación entera con el mismo modelo en la sección 11.4. Observe cómo el uso de las restricciones de elemento proporciona un modelo mucho más compacto y transparente.

Las restricciones *todas diferentes* y de *elemento* son sólo dos de las muchas restricciones globales disponibles (en la referencia seleccionada 6 se describen casi 40), pero ilustran muy bien el poder de la programación de restricciones para proporcionar un modelo compacto y fácil de leer de un problema complejo.

Investigación actual

La investigación actual relativa a la integración de la programación de restricciones y la programación entera se desarrolla en algunas trayectorias paralelas. Por ejemplo, el enfoque más directo es utilizar de manera simultánea un modelo de programación de restricciones y un modelo de programación entera para representar partes complementarias de un problema. Así, cada restricción relevante se incluye en cualquier modelo al que se ajuste o, cuando es factible, en ambos modelos. Como se aplican algoritmos de programación de restricciones y de programación entera a los modelos respectivos, la información fluye en ambos sentidos para enfocar la búsqueda de soluciones factibles (aquellas que satisfagan las restricciones de los dos modelos).

Este tipo de esquema de doble modelación se puede implantar con el lenguaje de programación para optimización (OPL, *Optimization Programming Language*) que se incorpora en el Sistema de Desarrollo ILOG OPL-CPLEX (ILOG es la compañía que produce el software de optimización CPLEX que se incluye en el OR Courseware). Después de emplear el lenguaje de modelación OPL, el Sistema de Desarrollo ILOG OPL-CPLEX puede invocar a un algoritmo de programación de restricciones (Optimizador ILOG CP) y a un solucionador de programación matemática (CPLEX) y después enviar informacion de uno al otro.

Aunque la doble programación es un buen primer paso, la meta es integrar por completo la programación de restricciones y la programación entera de manera que se pueda usar un solo modelo híbrido y un solo algoritmo. Cuando se logre este tipo de integración se podrá contar por completo con las fortalezas complementarias de las dos técnicas. A pesar de que el logro absoluto de esta meta permanece como un formidable reto de investigación, se continúa con buenos progresos en esta dirección. La referencia seleccionada 6 describe los avances más recientes en esta área.

Aún en esta etapa inicial ha habido un gran número de aplicaciones exitosas de la fusión de la programación matemática y la programación de restricciones. Las áreas de aplicación incluyen diseño de redes, determinación de rutas, rotación de tripulaciones, el problema clásico de transporte con costos lineales, administración de inventarios, gráficas por computadora, ingeniería de software, bases de datos, finanzas, ingeniería y optimización combinatoria, entre otras. Además, en la referencia seleccionada 2 se describe cómo las aplicaciones en la programación de actividades están demostrando ser un área particularmente fructífera para la aplicación de la programación de restricciones. Por ejemplo, debido a las múltiples y complicadas restricciones implicadas, en la actualidad se emplea la programación de restricciones para determinar el calendario regular de la Liga Nacional de Futbol (NFL) de Estados Unidos.

Estas aplicaciones apenas comienzan a cubrir el potencial de la integración de la programación de restricciones y la programación entera. El progreso subsecuente en la consecución de esta integración promete abrir muchas oportunidades nuevas e interesantes para aplicaciones importantes.

11.10 CONCLUSIONES

Los problemas de PE surgen con frecuencia cuando los valores de algunas o de todas las variables de decisión deben restringirse a valores enteros. Existen también muchas aplicaciones que necesitan decisiones sí o no (aquí se incluyen las relaciones combinatorias que pueden expresarse en términos de tales decisiones) que se puede representar por variables binarias (0-1). Estos factores han hecho que la programación entera sea una de las técnicas de IO de mayor aplicación.

Los problemas de PE son más difíciles de lo que serían sin la restricción de valores enteros, de manera que los algoritmos disponibles para programación entera, en general, son mucho menos eficientes que el método símplex. Sin embargo, ha habido enormes progresos en las dos últimas décadas en cuanto a la capacidad para resolver algunos (pero no todos) problemas enormes de IP con decenas y aun cientos de miles de variables enteras. Dicho progreso se ha debido a la combinación de tres factores: mejoras significativas en los algoritmos IP, mejoras en los algoritmos de programación lineal que se utilizan con los algoritmos IP y la gran velocidad de las computadoras. Sin embargo, los algoritmos IP también llegan en ocasiones a fallar en la resolución de problemas pequeños (aun en los que tienen solamente cien variables enteras). Diferentes características de un problema IP además de su tamaño, tienen una gran influencia en la velocidad a la que se puede resolver.

Los factores que determinan el tiempo de cálculo son el número de *variables enteras* y el hecho de que el problema tenga una *estructura especial* que pueda explotarse. En el caso de un número fijo de variables enteras, casi siempre es más fácil resolver problemas de PEB que problemas con variables enteras generales; pero agregar variables continuas (PEM) puede no significar un incremento sustancial del tiempo de cálculo. Para problemas especiales de PEB que contienen alguna estructura especial que puede aprovechar un *algoritmo especial*, es posible resolver problemas muy grandes (muy por encima de mil variables binarias) en forma rutinaria.

En la actualidad es común que se disponga de paquetes de computadora para algoritmos de PE en el software de programación matemática. Estos algoritmos casi siempre se basan en la técnica de *ramificación y acotamiento* o en alguna variación de ésta.

Hoy en día, los algoritmos más modernos de PE emplean el enfoque de *ramificación y corte*. Este enfoque algorítmico involucra la combinación del preprocesado automático del problema, la generación de cortes y las técnicas de ramificación y acotamiento. Continúa la investigación en esta área, junto con el desarrollo de paquetes de software elaborados para incorporar estas técnicas.

El último desarrollo en PE es comenzar a incorporar la *programación de restricciones*. En apariencia este enfoque expandirá de manera considerable la capacidad de formular y resolver problemas de PE.

En años recientes se han llevado a cabo muchas investigaciones con el fin de desarrollar algoritmos (incluso heurísticos) para programación entera *no lineal*, y en esta área la investigación continúa muy activa. (La referencia seleccionada 8 presenta los avances más actuales en esta área.)

REFERENCIAS SELECCIONADAS

1. Appa, G., L. Pitsoulis y H. P. Williams (eds.), *Handbook on Modeling for Discrete Optimization*, Springer, Nueva York, 2006.
2. Baptise, P., C. LePape y W. Nuijten, *Constraint-Based Scheduling:Applying Constraint Programming to Scheduling Problems*, Kluwer Academic Publishers (actualmente Springer), Boston, 2001.
3. Barnhart, C., P. Belobaba y A. R. Odoni, "Applications of Operations Research in the Air Transport Industry", en *Transportation Science*, **37**(4): 368-391, 2003.
4. Bixby, R. E., Z. Gu, E. Rothberg y R. Wunderling, "Mixed Integer Programming: A Progress Report", pp. 309-326 en M. Grötschel (ed.), *The Sharpest Cut: The Impact of Manfred Padberg and His Work*, MPS/SIAM Series en Optimización.
5. Hillier, F. S. y M. S. Hillier, *Introduction to Management Science: A Modeling and Case Studies Approach with Spreadsheets*, 3a. ed., McGraw-Hill/Irwin, Burr Ridge, IL, 2008, capítulo 7.
6. Hooker, J. N., *Integrated Methods for Optimization*, Springer, Nueva York, 2007.
7. Karlof, J. K., *Integer Programming: Theory and Practice*, CRC Press, Boca Raton, FL, 2006.
8. Li, D. y X. Sun, *Nolinear Integer Programming*, Springer, Nueva York, 2006.
9. Lübbecke, M. E. y J. Desrosiers: "Selected Topics in Column Generation", en *Operations Research*, **53**(6): 1007-1023, noviembre-diciembre de 2005.
10. Lustig, I. y J.-F. Puget, "Program Does Not Equal Program: Constraint Programming and its Relationship to Mathematical Programming", en *Interfaces*, **31**(6): 29-53, noviembre-diciembre de 2001.
11. Nemhauser, G. L., "Need and Potential for Real-Time Mixed Integer Programming", en *OR/MS Today*, **34**(1): 21-22, febrero de 2007.
12. Nemhauser, G. L. y L. A. Wolsey, *Integer and Combinatorial Optimization*, Wiley, Nueva York, 1988, reimpreso en 1999.

13. Schriver, A., *Theory of Linear and Integer Programming*, Wiley, Nueva York, 1986.
14. Williams, H. P., *Model Building in Mathematical Programming*, 4a. ed., Wiley, Nueva York, 1999.
15. Wolsey, L. A., *Integer Programming*, Wiley, Nueva York, 1998.
16. Wolsey, L. A., "Strong Formulations for Mixed Integer Programs: Valid Inequalities and Extended Formulations", en *Mathematical Programming Series B*, **97**(1-2): 423-447, 2003.

Algunas aplicaciones de programación entera ganadoras de premios

(En nuestro sitio en enternet —www.mhhe.com/hillier— se proporciona una liga hacia estos artículos.)

A1. Abbink, E., M. Fischetti, L. Kroon, G. Timmer y M. Vromans, "Reinventing Crew Scheduling at Netherlands Railways", en *Interfaces*, **35**(5): 393-401, septiembre-octubre de 2005.
A2. Armacost, A. P., C. Barnhart, K. A. Ware y A. M. Wilson, "UPS Optimizes Its Air Network", en *Interfaces*, **34**(1): 15-25, enero-febrero de 2004.
A3. Bertsimas, D., C. Darnell y R. Soucy, "Portfolio Construction Through Mixed-Integer Programming at Grantham, Mayo, Van Otterloo and Company", en *Interfaces*, **29**(1): 49-66, enero-febrero de 1999.
A4. Camm, J. D., T. E. Chorman, F. A. Dill, J. R. Evans, D. J. Sweeney y G. W. Wegryn, "Blending OR/MS, Judgment and GIS: Restructuring P&G's Supply Chain", en *Interfaces*, **27**(1): 128-142, enero-febrero de 1997.
A5. Denton, B. T., J. Forrest y R. J. Milne: "IBM Solves a Mixed-Integer Program to Optimize Its Semiconductor Supply Chain", en *Interfaces*, **36**(5): 386-399, septiembre-octubre de 2006.
A6. Gendron, B., "Scheduling Employees in Quebec's Liquor Stores with Integer Programming", en *Interfaces*, **35**(5): 402-410, septiembre-octubre de 2005.
A7. Gryffenberg, I., J. L. Lausberg, W. J. Smit, S. Uys, S. Botha, F. R. Hofmeyr, R. P. Nicolay, W. L. van der Merwe y G. J. Wessells, "Guns or Butter: Decision Support for Determinig the Size and Shape of the South African National Defense Force", en *Interfaces*, **27**(1): 7-28, enero-febrero de 1997.
A8. Martin, C., D. Jones y P. Keskinocak, "Optimizing On-Demand Aircraft Schedules for Fractional Aircraft Operators", en *Interfaces*, **33**(5): 22-35, septiembre-octubre de 2003.
A9. Metty, T., R. Harlan, Q. Samelson, T. Moore, T. Morris, R. Sorenson, A. Scneur, O. Raskina, R. Schneur, J. Kanner, K. Potts y J. Robbins, "Reinventing the Supplier Negotiation Process at Motorola", en *Interfaces*, **35**(1), 7-23, enero-febrero de 2005.
A10. Smith, B. C., R. Darrow, J. Elieson, D. Guenther, B. V. Rao y F. Zouaoui, "Travelocity Becomes a Travel Retailer", en *Interfaces*, **37**(1): 68-81, enero-febrero de 2007.
A11. Spencer III, T., A. J. Brigandi, D. R. Dargon y M. J. Sheehan: "AT&T's Telemarketing Site Selection System Offers Customer Support", en *Interfaces*, **20**(1): 83-96, enero-febrero de 1990.
A12. Subramanian, R., R. P. Scheff, Jr., J. D. Quillinan, D. S. Wiper y R. E. Marsten: "Coldstart: Fleet Assignment at Delta Air Lines", en *Interfaces*, **24**(1): 104-120, enero-febrero de 1994.
A13. Tyagi, R. y S. Bollapragada: "SES Americom Maximizes Satellite Revenues by Optimally Configuring Transponders", en *Interfaces* **33**(5): 36-44, septiembre-octubre de 2003.
A14. Yu, G., M. Argüello, G. Song, S. M. McCowan y A. White: "A New Era for Crew Recovery at Continental Airlines", en *Interfaces*, **33**(1): 5-22, enero-febrero de 2003.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos del capítulo 11

Ejemplo de demostración del OR Tutor:

Programación entera binaria, algoritmo de ramificación y acotamiento (*Binary Integer Programming Branch-and-Bound Algorithm*)
Programación entera mixta, algoritmo de ramificación y acotamiento (*Mixed Integer Programming Branch-and-Bound Algorithm*)

Rutinas interactivas del IOR Tutorial:

Introducción o revisión de un modelo general de programación lineal (*Enter or Revise a General Linear Programming Model*)
Solución interactiva de un programa entero binario (*Solve Binary Integer Program Interactively*)
Solución interactiva de un programa entero mixto (*Solve Mixed Integer Program Interactively*)

Complemento de Excel:

Premium Solver for Education

Archivos para resolver los ejemplos "Ch. 11—Integer Programming":

Archivos de Excel
Archivos de LINGO/LINDO
Archivos de MPL/CPLEX

Glosario para el capítulo 11:

Vea en el apéndice 1 la documentación del software.

PROBLEMAS

Los símbolos a la izquierda de algunos problemas (o de sus incisos) significan lo siguiente:

D: El ejemplo de demostración indicado que se mencionó en las ayudas de aprendizaje puede ser útil.
I: Se sugiere que utilice las rutinas interactivas correspondientes (la impresión registra su trabajo).
C: Use la computadora para resolver el problema; para ello utilice cualquier software disponible (o el que indique el profesor).

Un asterisco en el número del problema indica que al final del libro se proporciona al menos una respuesta parcial.

11.1-1. Reconsidere el ejemplo de la sección 11.1 de la California Manufacturing Co. El alcalde de San Diego se puso en contacto con el presidente de la compañía para persuadirlo de construir una fábrica y quizá un almacén en esa ciudad. Debido a los incentivos de impuestos que ofrece, el equipo del presidente estima que el valor presente neto de construir la planta sería de $7 millones y el capital que se requiere sería de $4 millones. El valor presente neto de construir el almacén sería de $5 millones y el capital requerido, de $3 millones. (Esta opción se considera sólo si se construye la planta en ese lugar.)

El presidente de la compañía quiere revisar el estudio de IO previo para incorporar estas nuevas alternativas al problema global. El objetivo todavía es encontrar la combinación factible de inversiones que maximice el valor presente neto total, dado que el capital disponible es de $10 millones.

***a*)** Formule un modelo de PEB para este problema.
***b*)** Muestre el modelo en una hoja de cálculo de Excel.
C ***c*)** Use la computadora para resolver este modelo.

11.1-2.* Una joven pareja, Eve y Steven, quiere dividir las principales tareas del hogar (ir de compras, cocinar, lavar platos y lavar ropa) entre los dos, de manera que cada uno tenga dos obligaciones y el tiempo total para hacer estas tareas sea mínimo. La eficiencia en cada una de las tareas difiere entre ellos; la siguiente tabla proporciona el tiempo que cada uno necesita para cada tarea:

	Horas necesarias por semana			
	Compras	**Cocinar**	**Lavar platos**	**Lavar ropa**
Eve	4.5 horas	7.8 horas	3.6 horas	2.9 horas
Steven	4.9 horas	7.2 horas	4.3 horas	3.1 horas

***a*)** Formule un modelo de PEB para este problema.
***b*)** Muestre el modelo en una hoja de cálculo de Excel.
C ***c*)** Use la computadora para resolver este modelo.

11.1-3. Una empresa de bienes raíces, Peterson & Johnson, analiza cinco proyectos de desarrollo posibles. La siguiente tabla muestra las ganancias a largo plazo estimadas (valor presente neto) que generaría cada proyecto y la inversión que se requiere para emprenderlo, en millones de dólares.

	Proyecto de desarrollo				
	1	**2**	**3**	**4**	**5**
Ganancia estimada	1	1.8	1.6	0.8	1.4
Capital requerido	6	12	10	4	8

Los propietarios de la empresa, Dave Peterson y Ron Johnson, reunieron $20 millones de capital de inversión para estos proyectos. Ellos quieren elegir la combinación de proyectos que maximice la ganancia total estimada a largo plazo (valor presente neto) sin invertir más de $20 millones.

***a*)** Formule un modelo de PEB para este problema.
***b*)** Muestre el modelo en una hoja de cálculo de Excel.
C ***c*)** Use la computadora para resolver este modelo.

11.1-4. El consejo directivo de General Wheeis Co., estudia seis grandes inversiones de capital. Cada inversión se puede hacer sólo una vez. Estas inversiones difieren en la ganancia estimada a largo

plazo (valor presente neto) que generarán, así como en la cantidad de capital que requiere cada uno, como se muestra en la siguiente tabla (en millones de dólares):

	Oportunidad de inversión					
	1	2	3	4	5	6
Ganancia estimada	15	12	16	18	9	11
Capital requerido	38	33	39	45	23	27

Se dispone de $100 millones de dólares como capital total para estas inversiones. Las oportunidades de inversión 1 y 2 son mutuamente excluyentes, lo mismo que 3 y 4. Más aún, la oportunidad 3 o la 4 no se pueden aprovechar a menos que se invierta en una de las dos primeras opciones. No existen restricciones de este tipo sobre las oportunidades de inversión 5 y 6. El objetivo es elegir la combinación de inversiones de capital que maximice la ganancia estimada a largo plazo (valor presente neto).

a) Formule el modelo de PEB para este problema.

C *b*) Use la computadora para resolver este modelo.

11.1-5. Reconsidere el problema 8.3-4, donde el entrenador de un equipo de natación debe asignar nadadores a los distintos equipos del combinado de 200 metros. Formule un modelo de PEB para este problema. Identifique los grupos de alternativas mutuamente excluyentes.

11.1-6. Vincent Cardoza es el propietario y director de un taller de maquinado que trabaja sobre pedido. El miércoles por la tarde recibió llamadas de dos clientes que necesitan órdenes urgentes. Un transportista de autos compactos necesita barras estabilizadoras. Una compañía de enganches para remolques requiere barras de remolque especiales para trabajo pesado. Ambos clientes quieren la mayor cantidad posible para el fin de semana (dos días hábiles). Como los dos productos usarán las mismas dos máquinas, Vincent debe decidir e informarles esta tarde cuántos productos de cada uno fabricará en los dos días siguientes.

Cada barra de remolque requiere 3.2 horas en la máquina 1 y 2 horas en la 2. Cada barra estabilizadora requiere 2.4 horas en la máquina 1 y 3 en la 2. La máquina 1 estará disponible 16 horas en los próximos dos días y la 2, 15 horas. La ganancia de cada barra de remolque producida será de $130 y la de cada barra estabilizadora será de $150.

Vincent quiere determinar la mezcla de estas cantidades de producción que maximizará su ganancia total.

a) Formule un modelo de PE para este problema.

b) Use un enfoque gráfico para resolver el modelo.

C *c*) Use la computadora para resolver este modelo.

11.1-7. Reconsidere el problema 8.2-21 de la contratista (Susan Meyer) que necesita acarrear grava de dos fosos a tres construcciones.

Ahora Susan necesita contratar camiones (y sus choferes) para el acarreo; cada camión puede usarse sólo para llevar grava de un foso a una construcción. Además de los costos de acarreo y grava que se especificaron, existe un costo fijo de $150 asociado a la contratación de cada camión. Un camión puede cargar 5 toneladas pero no tiene que ir lleno. Deben tomarse dos decisiones para cada combinación de foso y construcción, el número de camiones que se empleará y la cantidad de grava que van a acarrear.

a) Formule un modelo de PEM para este problema.

C *b*) Use la computadora para resolver el modelo.

11.2-1. Lea el artículo de referencia que describe en su totalidad el estudio de OR que se resume en el recuadro de aplicación que se presentó en la sección 11.2. Describa en forma breve la forma en que la programación entera se aplicó a este estudio. Después, elabore una lista de los diferentes beneficios financieros y no financieros que arrojó este estudio.

11.2-2. Seleccione una de las aplicaciones reales de PEB de las compañías u oficinas de gobierno que se mencionaron en la sección 11.2. Lea el artículo que la describe en la referencia de *Interfaces*. Escriba un resumen de dos páginas de la aplicación y sus beneficios.

11.2-3. Seleccione tres de las aplicaciones reales de PEB de compañías u oficinas de gobierno que se mencionó en la sección 11.2. Lea los artículos que las describen en las referencias de *Interfaces*. Escriba un resumen de una página de la aplicación y sus beneficios de cada uno.

11.3-1.* La división de investigación y desarrollo de la Progresive Company está en proceso de desarrollar cuatro líneas de posibles nuevos productos. La administración debe decidir cuáles de estos cuatro productos fabricar y a qué niveles. Ha pedido al departamento de IO que formule un modelo de programación matemática para encontrar la mezcla de productos más redituable.

La puesta en marcha de la fabricación de cualquier producto se asocia a un costo sustancial, que se proporciona en el primer renglón de la tabla. El objetivo de la administración es encontrar la mezcla de productos que maximice la ganancia total (ingreso neto total menos costos fijos).

	Producto			
	1	2	3	4
Costo de arranque	$50 000	$40 000	$70 000	$60 000
Ingreso marginal	$ 70	$ 60	$ 90	$ 80

Defina las variables de decisión continuas x_1, x_2, x_3 y x_4 como los niveles de producción de los productos 1, 2, 3 y 4. Por políticas de la empresa, la administración ha impuesto las siguientes restricciones sobre estas variables:

1. Como máximo, sólo deben fabricarse dos de estos productos.
2. Cualquiera de los productos 3 o 4 se puede producir sólo si se fabrica el producto 1 o el 2.
3. O bien $5x_1 + 3x_2 + 6x_3 + 4x_4 \leq 6\,000$
   o $4x_1 + 6x_2 + 3x_3 + 5x_4 \leq 6\,000.$

a) Introduzca variables binarias auxiliares para formular un modelo de PEB mixta de este problema.

C *b*) Use la computadora para resolver este modelo.

11.3-2. Suponga que un modelo matemático se ajusta a programación lineal excepto por la restricción $|x_1 - x_2| = 0$, o 3, o 6. Reformule esta restricción para ajustarla a PEM.

11.3-3. Suponga que un modelo matemático se ajusta a programación lineal excepto por las siguientes restricciones:

1. Al menos una de las dos desigualdades se cumple:

$$3x_1 - x_2 - x_3 + x_4 \leq 12$$
$$x_1 + x_2 + x_3 + x_4 \leq 15.$$

2. Al menos dos de las tres desigualdades se cumplen:

$$2x_1 + 5x_2 - x_3 + x_4 \leq 30$$
$$-x_1 + 3x_2 + 5x_3 + x_4 \leq 40$$
$$3x_1 - x_2 + 3x_3 - x_4 \leq 60.$$

Reformule estas restricciones para que se ajusten a PEM.

11.3-4. La compañía Toys-R-4-U ha desarrollado dos nuevos juguetes para su posible inclusión en la línea de productos la próxima temporada de Navidad. La preparación de instalaciones para iniciar la fabricación costaría $50 000 en el caso del juguete 1 y $80 000 en el del juguete 2. Una vez cubiertos estos costos, se obtendría una ganancia unitaria de $10 por el juguete 1 y $15 por el juguete 2.

La compañía tiene dos plantas que pueden producir estos juguetes. Sin embargo, para evitar la duplicidad de costos de preparación, sólo se usará una de ellas, y la elección depende de la maximización de la ganancia. Por razones administrativas, se usará la misma planta para ambos juguetes nuevos si se producen los dos.

El juguete 1 se puede producir a una tasa de 50 unidades por hora en la planta 1 y 40 por hora en la 2. El juguete 2 se puede producir a una tasa de 40 unidades por hora en la planta 1 y 25 por hora en la 2. Las plantas 1 y 2 tienen 500 y 700 horas de producción disponibles, respectivamente, antes de Navidad, que se pueden usar para producir estos juguetes.

No se sabe si estos juguetes continuarán fabricándose después de Navidad. Por lo tanto, el problema es determinar cuántas unidades (si se fabrican) de cada juguete nuevo deben producirse antes de Navidad a fin de maximizar la ganancia total.

***a*)** Formule un modelo de PEM para este problema.
C ***b*)** Use la computadora para resolver este modelo.

11.3-5.* La línea aérea Northeastern piensa comprar jets de pasajeros grandes, medianos y chicos. El precio de compra de cada avión grande será de $67 millones, $50 millones el de los medianos y $35 millones el de los chicos. El consejo directivo ha autorizado un compromiso máximo de $1.5 mil millones para realizar estas compras. Sin que importe qué aviones se compren, se espera que las distancias de los trayectos sean lo suficientemente grandes como para que los aviones se utilicen, en esencia, a su capacidad máxima. Se estima que la ganancia neta anual (después de restar los costos de recuperación de capital) de un avión grande será de $4.2 millones, $3 millones si se trata de un avión mediano y $2.3 millones de cada avión chico.

Se piensa que la compañía podrá disponer de suficientes pilotos entrenados para operar 30 aviones nuevos. Si sólo se compraran aviones chicos, las instalaciones de mantenimiento podrían manejar 40 aviones, pero cada avión mediano equivale a aviones chicos y cada avión grande equivale a aviones chicos, en términos de la utilización de las instalaciones de mantenimiento.

Esta información se obtuvo en un análisis preliminar del problema. Más adelante se llevará a cabo un estudio más detallado. Si se toman estos datos como una primera aproximación, la gerencia desea saber cuántos aviones de cada tipo debe comprar a fin de maximizar la ganancia.

***a*)** Formule un modelo de PE para representar este problema.
C ***b*)** Use la computadora para resolverlo.
***c*)** Utilice una representación binaria de las variables para reformular el modelo de PE del inciso *a*) como un modelo de PEB.
C ***d*)** Utilice la computadora para resolver el modelo de PEB del inciso *c*). Después use esta solución óptima para identificar una solución óptima para el modelo de PE que se formuló en el inciso *a*).

11.3-6. Considere el ejemplo de PE de dos variables de la sección 11.5, que se presentó en la figura 11.3.

***a*)** Use una representación binaria de variables para reformular este modelo como un problema de PEB.
C ***b*)** Utilice la computadora para resolver este problema de PEB. Después use la solución óptima para identificar una solución óptima para el modelo original de PE.

11.3-7. La compañía aérea Fly-Right construye jets pequeños que vende a corporaciones para uso ejecutivo. Para cumplir con sus necesidades, en ocasiones los clientes ordenan aviones con diseño especial. Cuando es así, se incurre en un costo de preparación para iniciar la producción de las aeronaves.

Fly-Right acaba de recibir pedidos de tres clientes con fechas de entrega cercanas. Debido a que las instalaciones de producción están comprometidas para cumplir contratos anteriores, no podrán aceptar los tres pedidos. En consecuencia, debe decidirse el número de aviones que producirán (si lo hacen) para cada uno de los tres clientes.

Los datos relevantes se presentan en la siguiente tabla. El primer renglón contiene los costos fijos para iniciar la producción de aviones de cada cliente. Con la producción en marcha, el ingreso neto marginal (precio de compra menos costo marginal de producción) de cada avión se presenta en el segundo renglón. El tercero contiene los porcentajes de capacidad de producción disponibles para cada avión. El último renglón indica el número máximo de aviones pedidos por cada cliente (pero aceptarían menos).

	Cliente		
	1	2	3
Costo de arranque	$3 millones	$2 millones	0
Ingreso neto marginal	$2 millones	$3 millones	$0.8 millones
Capacidad usada/avión	20%	40%	20%
Pedido máximo	3 aviones	2 aviones	5 aviones

Fly-Right desea determinar cuántos aviones debe producir para cada cliente (si lo hace) de modo que se maximice su ganancia total (ingresos netos menos costos fijos).

***a*)** Formule un modelo con variables enteras y variables binarias para representar este problema.
C ***b*)** Use la computadora para resolver el modelo.

11.4-1. Reconsidere el problema 11.3-7 de Fly-Right. Un análisis más detallado de los factores de costo e ingreso indica que la ganancia potencial de producir los aviones de cada cliente no se puede expresar sólo en términos del *costo fijo* y un *ingreso neto marginal* por avión producido. Más bien, las ganancias se proporcionan en la siguiente tabla.

Aviones producidos	Ganancias por cliente		
	1	2	3
0	0	0	0
1	−$1 millón	$1 millón	$1 millón
2	$2 millones	$5 millones	$3 millones
3	$4 millones		$5 millones
4			$6 millones
5			$7 millones

***a*)** Formule un modelo de PEB para manejar este problema que incluya restricciones para las *alternativas mutuamente excluyentes*.

C ***b*)** Use la computadora para resolver el modelo formulado en el inciso *a*). Después, utilice esta solución óptima para identificar el número óptimo de aviones que se debe fabricar para cada cliente.

***c*)** Formule otro modelo de PEB para representar este problema que incluya restricciones sobre las *decisiones contingentes*.

C ***d*)** Repita el inciso *b*) para el modelo formulado en el inciso *c*).

11.4-2. Reconsidere el problema de la Wyndor Glass Co., que se presentó en la sección 3.1. La administración ha decidido que sólo debe fabricarse uno de los dos nuevos productos y la elección debe hacerse para maximizar la ganancia. Introduzca *variables binarias auxiliares* para formular un modelo de PEM para construir esta nueva versión del problema.

11.4-3.* Reconsidere el problema 3.1-11, en el cual la administración de Omega estudia la posibilidad de dedicar el exceso de capacidad a uno o más de tres productos. (Vea información adicional de este problema en las respuestas al final del libro.) Ahora agregue la restricción de la administración de que no deben producirse más de dos de los tres nuevos productos.

***a*)** Introduzca *variables binarias auxiliares* para formular un modelo de PEM para presentar esta nueva versión del problema.

C ***b*)** Use la computadora para resolver el modelo.

11.4-4. Considere el siguiente problema de programación entera no lineal.

$$\text{Maximizar} \qquad Z = 4x_1^2 - x_1^3 + 10x_2^2 - x_2^4,$$

sujeta a

$$x_1 + x_2 \leq 3$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0$$

x_1 y x_2 son enteros.

Este problema se puede reformular de dos maneras como un problema de PEB equivalente (con una función objetivo lineal) y seis variables binarias (y_{1j} y y_{2j} para $j = 1, 2, 3$), según la interpretación que se dé a las variables binarias.

***a*)** Formule un modelo de PEB que represente este problema donde las variables binarias tienen la siguiente interpretación,

$$y_{ij} = \begin{cases} 1 & \text{si } x_i = j \\ 0 & \text{de otra manera.} \end{cases}$$

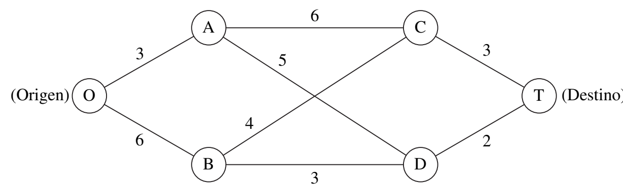
C ***b*)** Utilice la computadora para resolver el modelo que se formuló en el inciso *a*), y después encuentre una solución óptima para (x_1, x_2) en el problema original.

***c*)** Formule un modelo de PEB para este problema donde las variables binarias tengan la siguiente interpretación:

$$y_{ij} = \begin{cases} 1 & \text{si } x_i \geq j \\ 0 & \text{de otra manera.} \end{cases}$$

C ***d*)** Use la computadora para resolver el modelo que se formuló en el inciso *c*), y después encuentre una solución óptima para (x_1, x_2) en el problema original.

11.4-5.* Considere el siguiente tipo especial de *problema de la ruta más corta* (sección 9.3) en el que los nodos están en columna y las únicas trayectorias posibles se mueven siempre hacia adelante, una columna a la vez.



Los números colocados junto a las trayectorias son distancias y el objetivo es encontrar la ruta más corta del origen al destino.

Este problema también se puede formular como un modelo de PEB que abarca tanto alternativas mutuamente excluyentes como decisiones contingentes.

***a*)** Formule este modelo. Identifique las restricciones de las alternativas mutuamente excluyentes y las de decisiones contingentes.

C ***b*)** Use la computadora para resolver este problema.

11.4-6. Speedy Delivery proporciona un servicio que entrega paquetes grandes en dos días, en todo Estados Unidos. Cada mañana se cargan los paquetes que llegaron a cada centro de recolección durante la noche en los camiones de reparto para su entrega en el área. En razón de que la competencia en este negocio se basa en la rapidez de la entrega, los paquetes se dividen según sus destinos geográficos de manera que se minimice el tiempo promedio necesario para realizar las entregas.

Esta mañana, la despachadora del centro de recolección de Blue River Valley, Sharon Lofton, tiene mucho trabajo. Sus tres choferes llegarán en menos de una hora para el reparto. Hay nueve paquetes que entregar en lugares muy alejados entre sí. Como siempre, Sharon introduce estos lugares en la computadora para usar Dispatcher, el software especial del sistema de apoyo. El programa usa las ubicaciones para generar un buen número de rutas posibles para cada camión. Estas rutas se muestran en la siguiente tabla (donde los números en cada columna indican el orden de las entregas), junto con los tiempos que se requieren para el recorrido.

Ubicación de la entrega	Rutas atractivas posibles									
	1	2	3	4	5	6	7	8	9	10
A	1				1				1	
B		2		1		2			2	2
C			3	3			3		3	
D	2					1		1		
E			2	2		3				
F		1			2					
G	3						1	2		3
H			1		3					1
I		3		4			2			
Tiempo (en horas)	6	4	7	5	4	6	5	3	7	6

Dispatcher es un sistema interactivo que muestra estas rutas para que Sharon las apruebe o modifique. (Quizá la computadora no sepa que una inundación ha hecho que una ruta sea no factible.) Si Sharon aprueba las rutas como posibilidades atractivas con tiempos estimados razonables, el programa formula y resuelve un modelo de PEB para elegir las tres rutas que minimizan el tiempo total e incluye cada lugar de entrega sólo en una ruta. Esta mañana, Sharon aprueba todas las rutas.

a) Formule el modelo de PEB.

C ***b***) Use la computadora para resolver el problema.

11.4-7. Cada vez más estadounidenses se mudan a un clima más templado cuando se retiran. Para aprovechar esta tendencia, Sunny Skies Unlimited emprenderá un proyecto de desarrollo urbano. Se creará una comunidad nueva por completo (llamada Pilgrim Heaven) que cubrirá varias millas cuadradas. Una de las decisiones que se debe tomar es la ubicación más conveniente de las dos estaciones de bomberos que se le asignaron a la comunidad. Para propósitos de planeación, se dividió Pilgrim Heaven en cinco sectores, con sólo una estación de bomberos en cada uno de ellos. Cada estación debe responder a *todos* los llamados que reciba del sector en el que se localiza y las de otros que se le asignen. Entonces, las decisiones son: 1) los sectores que albergarán una estación de bomberos y 2) la asignación de cada uno de los otros sectores a una de las estaciones. El objetivo es minimizar el promedio global de los *tiempos de respuesta* a los incendios.

La siguiente tabla proporciona el tiempo promedio de respuesta (en minutos) a un incendio en cada sector (columnas) si el servicio se presta desde la estación de un sector dado (renglones). El último renglón proporciona el pronóstico del número promedio de incendios diarios que ocurrirán en cada uno de los sectores.

Estación asignada localizada en el sector	Tiempos de respuesta en minutos Incendios en el sector				
	1	2	3	4	5
1	5	12	30	20	15
2	20	4	15	10	25
3	15	20	6	15	12
4	25	15	25	4	10
5	10	25	15	12	5
Frecuencia de emergencia (por día)	2 por día	1 por día	3 por día	1 por día	3 por día

Formule un modelo completo de PEB para representar el problema. Identifique la restricción de las alternativas mutuamente excluyentes o de las decisiones contingentes.

11.4-8. Reconsidere el problema 11.4-7. Ahora, la administración de Sunny Skies desea que la decisión sobre el lugar en donde se ubicarán las estaciones de bomberos se base en los costos.

El costo de asignar una estación de bomberos en el sector 1 es de \$300 000; de \$350 000 en el 2; de \$600 000 en el 3; de \$450 000 en el 4 y de \$700 000 en el 5. Ahora, el nuevo objetivo de la gerencia es:

> Determinar cuáles sectores deben tener una estación para minimizar el costo total de las estaciones al mismo tiempo que se asegure que cada sector tenga al menos una estación lo suficientemente cerca para responder a un incendio en no más de 12 minutos (en promedio).

Observe que al contrario del problema original, el número total de estaciones de bomberos no es fijo. Aún más, si un sector sin estación tiene más de una estación a 12 minutos o menos, ya no es necesario asignar este sector a sólo una de las estaciones.

a) Formule un modelo de PEB pura con cinco variables binarias para representar este problema.

b) ¿Éste es un *problema de cobertura de conjunto*? Explique e identifique los conjuntos relevantes.

C ***c***) Use la computadora para resolver el modelo formulado en *a*).

11.4-9. Suponga que un estado envía R personas a la cámara de diputados. Existen D municipios en el estado ($D > R$) y la legislatura estatal quiere agrupar estos municipios en R distritos electorales, cada uno de los cuales elegirá un diputado. La población total del estado es P, y la legislatura desea formar distritos cuya población se aproxime a $p = P/R$. Suponga que el comité legislativo que estudia el problema ha generado una larga lista de N *candidatos* a ser convertidos en distritos ($N > R$). Cada candidato abarca municipios contiguos y una población total p_j ($j = 1, 2, \dots, N$) aceptablemente cercana a p. Defina $c_j = |p_j - p|$. Cada municipio i ($i = 1, 2, \dots, D$) está incluido al menos con un candidato a ser distrito y casi en todos los casos estará incluido con un gran número de candidatos (con el fin de proporcionar muchas maneras factibles de seleccionar un conjunto de R candidatos que incluyan a cada municipio exactamente una vez). Defina

$$a_{ij} = \begin{cases} 1 & \text{si el municipio } i \text{ se incluye con el candidato } j \\ 0 & \text{si no.} \end{cases}$$

Dados los valores de las c_j y las a_{ij} el objetivo es seleccionar R de estos N posibles distritos, tales que cada municipio esté contenido en un solo distrito y que la más grande de las c_j asociadas sea lo más pequeña posible.

Formule un modelo de PEB para este problema.

11.5-1. Lea el artículo de referencia que describe en su totalidad el estudio de OR que se resume en el recuadro de aplicación que se presentó en la seccion 11.5. Describa brevemente la forma en que se aplicó la programación entera en el estudio. Después, elabore una lista de los beneficios financieros y de otro tipo que arrojó este estudio.

11.5-2.* Considere el siguiente problema de PE.

$$\text{Maximizar} \quad Z = 5x_1 + x_2,$$

sujeta a

$$\begin{aligned} -x_1 + 2x_2 &\leq 4 \\ x_1 - x_2 &\leq 1 \\ 4x_1 + x_2 &\leq 12 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0$$
$$x_1, x_2 \text{ son enteros.}$$

a) Resuelva este problema de manera gráfica.

b) Resuelva el relajamiento de PL de manera gráfica. Redondee esta solución a la solución entera *más cercana* y verifique si es factible. Después enumere *todas* las soluciones redondeadas. Para ello redondee esta solución del relajamiento de PL de *todas* las maneras posibles (es decir, redondee cada valor no entero tanto

hacia arriba como hacia abajo). Verifique la factibilidad de cada solución redondeada y, si es factible, calcule Z. ¿Alguna de estas soluciones redondeadas es óptima para el problema de PE?

11.5-3. Siga las instrucciones del problema 11.5-2 para resolver el siguiente problema de PE.

Maximizar $Z = 220x_1 + 80x_2$,

sujeta a

$$\begin{aligned} 5x_1 + 2x_2 &\leq 16 \\ 2x_1 - x_2 &\leq 4 \\ -x_1 + 2x_2 &\leq 4 \end{aligned}$$

y

$x_1 \geq 0, \quad x_2 \geq 0$

x_1, x_2 son enteros.

11.5-4. Siga las instrucciones del problema 11.5-2 para solucionar el siguiente problema de PEB.

Maximizar $Z = 10x_1 + 25x_2$,

sujeta a

$$\begin{aligned} 19x_1 + 6x_2 &\leq 15 \\ 5x_1 - 15x_2 &\leq 15 \end{aligned}$$

y

x_1, x_2 son binarios.

11.5-5. Siga las instrucciones del problema 11.5-2 para solucionar el siguiente problema de PE.

Maximizar $Z = -5x_1 + 25x_2$,

sujeta a

$$\begin{aligned} -3x_1 + 30x_2 &\leq 27 \\ 3x_1 + x_2 &\leq 4 \end{aligned}$$

y

x_1, x_2 son binarios.

11.5-6. Diga si cada una de las afirmaciones siguientes es falsa o verdadera y justifique su respuesta con referencia al texto específico (con la cita de página) del capítulo.

***a*)** En general es más fácil resolver problemas de programación lineal que problemas de PE.

***b*)** Por lo general, en los problemas de PE el número de variables enteras es más importante para determinar la dificultad en los cálculos que el número de restricciones funcionales.

***c*)** Para resolver un problema de PE con un procedimiento aproximado se puede aplicar el método símplex al relajamiento de PE y redondear cada valor no entero al entero más cercano. El resultado será factible pero no necesariamente la solución óptima del problema de PE.

D,I **11.6-1.*** Utilice el algoritmo de ramificación y acotamiento de PEB de la sección 11.6 para resolver el siguiente problema de manera interactiva.

Maximizar $Z = 2x_1 - x_2 + 5x_3 - 3x_4 + 4x_5$,

sujeta a

$$\begin{aligned} 3x_1 - 2x_2 + 7x_3 - 5x_4 + 4x_5 &\leq 6 \\ x_1 - x_2 + 2x_3 - 4x_4 + 2x_5 &\leq 0 \end{aligned}$$

y

x_j es binario, para $j = 1, 2, \ldots, 5$.

D,I **11.6-2.** Aplique el algoritmo de ramificación y acotamiento de PEB de la sección 11.6 para resolver el siguiente problema de manera interactiva.

Minimizar $Z = 5x_1 + 6x_2 + 7x_3 + 8x_4 + 9x_5$,

sujeta a

$$\begin{aligned} 3x_1 - x_2 + x_3 + x_4 - 2x_5 &\geq 2 \\ x_1 + 3x_2 - x_3 - 2x_4 + x_5 &\geq 0 \\ -x_1 - x_2 + 3x_3 + x_4 + x_5 &\geq 1 \end{aligned}$$

y

x_j es binario, para $j = 1, 2, \ldots, 5$.

D,I **11.6-3.** Utilice el algoritmo de ramificación y acotamiento de PEB de la sección 11.6 para resolver el siguiente problema de manera interactiva.

Maximizar $Z = 3x_1 + 3x_2 + 5x_3 - 2x_4 - x_5$,

sujeta a

$$\begin{aligned} x_1 + 2x_2 - 3x_4 - x_5 &\leq 0 \\ -15x_1 + 30x_2 - 35x_3 + 45x_4 + 45x_5 &\geq 50 \end{aligned}$$

y

x_j es binario, para $j = 1, 2, \ldots, 5$.

D,I **11.6-4.** Reconsidere el problema 11.3-6*a*). Use el algoritmo de ramificación y acotamiento de PEB de la sección 11.6 para resolver este modelo de manera interactiva.

D,I **11.6-5.** Reconsidere el problema 11.4-8*a*). Utilice el algoritmo de PEB de la sección 11.6 para resolver este problema de manera interactiva.

11.6-6. Considere las siguientes afirmaciones sobre un problema de PE pura (en la forma de maximización) y su relajamiento de PE. Diga si cada una es falsa o verdadera y justifique su respuesta.

***a*)** La región factible del relajamiento de PE es un subconjunto de la región factible del problema de PE.

***b*)** Si una solución óptima del relajamiento de PE es una solución entera, entonces el valor óptimo de la función objetivo es el mismo en ambos problemas.

***c*)** Si una solución no entera es factible para el relajamiento de PE, entonces la solución entera más cercana (al redondear cada variable al entero más próximo) es una solución factible para el problema de PE.

11.6-7.* Considere el problema de asignación que tiene la siguiente tabla de costos:

		Tarea				
		1	2	3	4	5
	1	39	65	69	66	57
	2	64	84	24	92	22
Asignado	3	49	50	61	31	45
	4	48	45	55	23	50
	5	59	34	30	34	18

a) Diseñe un algoritmo de ramificación y acotamiento para resolver problemas de asignación; especifique cómo se deben realizar los pasos de ramificación, acotamiento y sondeo. (*Sugerencia:* Para los asignados que todavía no se incluyen en el subproblema actual, forme el relajamiento de PL mediante la eliminación de las restricciones de que cada uno debe realizar exactamente una tarea.)

b) Utilice este algoritmo para resolver el problema.

11.6-8. Es necesario realizar cinco trabajos en un cierto número de máquinas. El tiempo de preparación de estos trabajos depende del trabajo que les antecede, según se muestra en la siguiente tabla:

		Tiempo de preparación				
		Trabajo				
		1	2	3	4	5
	Ninguno	4	5	8	9	4
	1	—	7	12	10	9
Trabajo precedente	2	6	—	10	14	11
inmediato	3	10	11	—	12	10
	4	7	8	15	—	7
	5	12	9	8	16	—

El objetivo es programar la *secuencia* de trabajos que minimice la suma de los tiempos de preparación resultantes.

a) Diseñe un algoritmo de ramificación y acotamiento para problemas de secuenciación; especifique los pasos de ramificación, de acotamiento y de sondeo.

b) Use este algoritmo para resolver el problema.

11.6-9.* Considere el siguiente problema de PEB *no lineal*.

$$\text{Maximizar} \quad Z = 80x_1 + 60x_2 + 40x_3 + 20x_4 - (7x_1 + 5x_2 + 3x_3 + 2x_4)^2,$$

sujeta a

x_j es binario, para $j = 1, 2, 3, 4$.

Dados los valores de las primeras k variables $x_1, \ldots, x_k$, donde $k = 0$, 1, 2 o 3, una cota superior del valor de Z que se puede lograr mediante las soluciones factibles es

$$\sum_{j=1}^{k} c_j x_j - \left(\sum_{j=1}^{k} d_j x_j\right)^2 + \sum_{j=k+1}^{4} \text{máx}\left\{0,\ c_j - \left[\left(\sum_{i=1}^{k} d_i x_i + d_j\right)^2 - \left(\sum_{i=1}^{k} d_i x_i\right)^2\right]\right\},$$

donde $c_1 = 80$, $c_2 = 60$, $c_3 = 40$, $c_4 = 20$, $d_1 = 7$, $d_2 = 5$, $d_3 = 3$, $d_4 = 2$. Utilice esta cota para resolver el problema mediante la técnica de ramificación y acotamiento.

11.6-10. Considere el relajamiento de Lagrange descrito casi al final de la sección 11.6.

a) Si **x** es una solución factible para un problema de PEM, demuestre que **x** también debe ser factible para el relajamiento de Lagrange correspondiente.

b) Si **x*** es una solución óptima para un problema de PEM, con Z como valor de la función objetivo, demuestre que $Z \leq Z_R^*$, donde Z_R^* es el valor óptimo de la función objetivo del relajamiento de Lagrange correspondiente.

11.7-1. Lea el artículo de referencia que describe en su totalidad el estudio de OR que se resume en el Recuadro de aplicación que se presentó en la seccion 11.7. Describa brevemente la forma en que se aplicó la programación entera en el estudio. Después, elabore una lista de los beneficios financieros y no financieros que arrojó este estudio.

11.7-2.* Considere el siguiente problema de PE.

$$\text{Maximizar} \quad Z = -3x_1 + 5x_2,$$

sujeta a

$$5x_1 - 7x_2 \geq 3$$

y

$x_j \leq 3$
$x_j \geq 0$
x_j es entero, para $j = 1, 2$.

a) Resuelva este problema de manera gráfica.

b) Utilice el algoritmo de ramificación y acotamiento de PEM que se presentó en la sección 11.7 para resolver este problema en forma manual. Resuelva el relajamiento de PL de cada subproblema por el método *gráfico*.

c) Utilice la representación binaria de variables enteras para reformular este problema como uno de PEB.

D,I ***d***) Aplique el algoritmo de ramificación y acotamiento de PEB de la sección 11.6 para resolver el inciso *c*) del modelo.

11.7-3. Siga las instrucciones del problema 11.7-2 para resolver el siguiente modelo de PE.

$$\text{Minimizar} \quad Z = 15x_1 + 10x_2,$$

sujeta a

$$15x_1 + 5x_2 \geq 30$$
$$10x_1 + 10x_2 \geq 30$$

y

$x_1 \geq 0, \quad x_2 \geq 0$
x_1, x_2 son enteros.

11.7-4. Reconsidere el problema 11.5-2.

a) Use el algoritmo de ramificación y acotamiento de PEM de la sección 11.7 para resolver este problema en forma manual. Resuelva el relajamiento de PL de cada subproblema por el método *gráfico*.

D,I ***b***) Utilice la rutina interactiva para este algoritmo del OR Courseware para resolver el problema.

C ***c***) Verifique su respuesta mediante el uso de la rutina automática para resolver el problema.

D,I **11.7-5.** Considere el ejemplo de PE de la sección 11.5 que se presentó en la figura 11.3. Utilice el algoritmo de ramificación y acotamiento de PEM de la sección 11.7 para resolver este problema de manera interactiva.

D,I **11.7-6.** Reconsidere el problema 11.3-5*a*). Use el algoritmo de ramificación y acotamiento de la sección 11.7 para resolver este problema de PE de manera interactiva.

11.7-7. Un taller de producción intermitente fabrica dos productos. Cada unidad del primer producto requiere 3 horas en la máquina 1 y 2 horas en la máquina 2. Cada unidad del segundo producto requiere 2 horas en la máquina 1 y 3 horas en la 2. La máquina 1 está disponible 8 horas al día y la 2 sólo 7 horas al día. La ganancia por unidad vendida del primer producto es de 16 y la del segundo, de 10. La cantidad diaria que se fabrica de cada producto debe ser un múltiplo entero de 0.25. El objetivo es determinar la mezcla de las cantidades de producción que maximizará la ganancia.

***a*)** Formule un modelo de PE para resolver este problema.

***b*)** Resuelva este modelo de manera gráfica.

***c*)** Utilice el análisis gráfico para aplicar el algoritmo de ramificación y acotamiento de PEM de la sección 11.7 para resolver el modelo.

D,I ***d*)** Aplique el algoritmo interactivo del IOR Tutorial para resolver el modelo.

C ***e*)** Verifique sus respuestas de los incisos *b*), *c*) y *d*) con una rutina automática para resolver el modelo.

D,I **11.7-8.** Use el algoritmo de ramificación y acotamiento de PEM de la sección 11.7 para resolver el siguiente problema de PEM de manera interactiva.

Maximizar $Z = 20x_1 + 10x_2 + 25x_3 + 20x_4$,

sujeta a

$$\begin{aligned} x_1 + x_2 + x_3 + 2x_4 &\le 12 \\ 3x_1 + x_2 + 2x_3 + 2x_4 &\le 20 \\ x_1 + 2x_2 + 5x_3 + 3x_4 &\le 30 \end{aligned}$$

y

$x_j \ge 0$, para $j = 1, 2, 3, 4$

x_j es entero, para $j = 1, 2, 3$.

D,I **11.7-9.** Utilice el algoritmo de ramificación y acotamiento de PEM de la sección 11.7 para resolver el siguiente problema en forma interactiva.

Maximizar $Z = 3x_1 + 4x_2 + 2x_3 + x_4 + 2x_5$,

sujeta a

$$\begin{aligned} 2x_1 - x_2 + x_3 + x_4 + x_5 &\le 3 \\ -x_1 + 3x_2 + x_3 - x_4 - 2x_5 &\le 2 \\ 2x_1 + x_2 - x_3 + x_4 + 3x_5 &\le 1 \end{aligned}$$

y

$x_j \ge 0$, para $j = 1, 2, 3, 4, 5$

x_j es binario, para $j = 1, 2, 3$.

D,I **11.7-10.** Utilice el algoritmo de ramificación y acotamiento de PEM de la sección 11.7 para resolver el siguiente problema de manera interactiva.

Minimizar $Z = 5x_1 + x_2 + x_3 + 2x_4 + 3x_5$,

sujeta a

$$\begin{aligned} x_2 - 5x_3 + x_4 + 2x_5 &\ge -2 \\ 5x_1 - x_2 + x_5 &\ge 7 \\ x_1 + x_2 + 6x_3 + x_4 &\ge 4 \end{aligned}$$

y

$x_j \ge 0$, para $j = 1, 2, 3, 4, 5$

x_j es entero, para $j = 1, 2, 3$.

11.8-1.* Utilice cada una de las siguientes restricciones de PEB pura para fijar tantas variables como sea posible.

***a*)** $4x_1 + x_2 + 3x_3 + 2x_4 \le 2$

***b*)** $4x_1 - x_2 + 3x_3 + 2x_4 \le 2$

***c*)** $4x_1 - x_2 + 3x_3 + 2x_4 \ge 7$

11.8-2. Utilice cada una de las siguientes restricciones de PEB pura para fijar tantas variables como sea posible.

***a*)** $20x_1 - 7x_2 + 5x_3 \le 10$

***b*)** $10x_1 - 7x_2 + 5x_3 \ge 10$

***c*)** $10x_1 - 7x_2 + 5x_3 \le -1$

11.8-3. Use el conjunto de restricciones para el *mismo* problema de PEB pura para fijar tantas variables como sea posible. También identifique las restricciones que se convierten en redundantes al fijar una variable.

$$\begin{aligned} 3x_3 - x_5 + x_7 &\le 1 \\ x_2 + x_4 + x_6 &\le 1 \\ x_1 - 2x_5 + 2x_6 &\ge 2 \\ x_1 + x_2 - x_4 &\le 0 \end{aligned}$$

11.8-4. Identifique cuáles restricciones de problemas de PE pura se convierten en redundantes debido a las restricciones binarias. Explique por qué son o no redundantes.

***a*)** $2x_1 + x_2 + 2x_3 \le 5$

***b*)** $3x_1 - 4x_2 + 5x_3 \le 5$

***c*)** $x_1 + x_2 + x_3 \ge 2$

***d*)** $3x_1 - x_2 - 2x_3 \ge -4$

11.8-5. En la sección 11.8, al final de la subsección sobre estrechar las restricciones, se indicó que la restricción $4x_1 - 3x_2 + x_3 + 2x_4 \le 5$ se pueden estrechar a $2x_1 - 3x_2 + x_3 + 2x_4 \le 3$ y después a $2x_1 - 2x_2 + x_3 + 2x_4 \le 3$. Use el procedimiento para estrechar las restricciones para confirmar los resultados.

11.8-6. Aplique el procedimiento para *estrechar las restricciones* a la restricción de un problema de PEB pura.

$$5x_1 - 10x_2 + 15x_3 \le 15.$$

11.8-7. Aplique el procedimiento para *estrechar las restricciones* a la restricción de un problema de PEB pura:

$$x_1 - x_2 + 3x_3 + 4x_4 \ge 1.$$

11.8-8. Aplique el procedimiento para *estrechar las restricciones* a las restricciones de un problema de PEB pura:

***a*)** $x_1 + 3x_2 - 4x_3 \le 2$.

***b*)** $3x_1 - x_2 + 4x_3 \ge 1$.

11.8-9. En la sección 11.8 se utilizó un ejemplo de PEB pura con la restricción $2x_1 + 3x_2 \le 4$ para ilustrar cómo estrechar restriccio-

nes. Demuestre que al aplicar el procedimiento para generar planos cortantes a esta restricción se llega a la misma nueva restricción $x_1 + x_2 \leq 1$.

11.8-10. Una de las restricciones de cierto problema de PEB pura es

$$x_1 + 3x_2 + 2x_3 + 4x_4 \leq 5.$$

Identifique todas sus cubiertas mínimas y proporcione los planos cortantes correspondientes.

11.8-11. Una de las restricciones de cierto problema de PEB pura es

$$25x_1 + 15x_2 + 20x_3 + 10x_4 \leq 35.$$

Identifique todas las cubiertas mínimas para ella y proporcione los planos cortantes correspondientes.

11.8-12. Genere tantos planos cortantes como sea posible a partir de la siguiente restricción de un problema de PEB pura.

$$3x_1 + 5x_2 + 4x_3 + 8x_4 \leq 10.$$

11.8-13. Genere tantos planos cortantes como sea posible a partir de la siguiente restricción de un problema de PEB pura:

$$5x_1 + 3x_2 + 7x_3 + 4x_4 + 6x_5 \leq 9.$$

11.8-14. Considere el siguiente problema de PEB pura.

$$\text{Maximizar} \quad Z = 2x_1 + 3x_2 + x_3 + 4x_4 + 3x_5 + 2x_6 + 2x_7 + x_8 + 3x_9,$$

sujeta a

$$\begin{aligned} 3x_2 + x_4 + x_5 &\geq 3 \\ x_1 + x_2 &\leq 1 \\ x_2 + x_4 - x_5 - x_6 &\leq -1 \\ x_2 + 2x_6 + 3x_7 + x_8 + 2x_9 &\geq 4 \\ -x_3 + 2x_5 + x_6 + 2x_7 - 2x_8 + x_9 &\leq 5 \end{aligned}$$

y

todo x_j es binario.

Desarrolle la formulación más estrecha posible por medio de las técnicas del procesamiento automático del problema (fijar variables, eliminar restricciones redundantes y estrechar restricciones). Utilice la formulación estrecha para determinar una solución óptima por inspección.

11.9-1. Considere el siguiente problema.

$$\text{Maximizar } Z = 10x_1 + 30x_2 + 40x_3 + 30x_4,$$

sujeta a

$x_1 \in \{2, 3\},\ x_2 \in \{2, 4\},\ x_3 \in \{3, 4\},\ x_4 \in \{1, 2, 3, 4\}$,
todas estas variables deben tener valores diferentes,
$x_1 + x_2 + x_3 + x_4 \leq 10$.

Use las técnicas de programación de restricciones (reducción del dominio, propagación de restricciones, procedimiento de búsqueda y enumeración) para identificar todas las soluciones factibles y después encontrar una solución óptima. Muestre su trabajo.

11.9-2. Considere el siguiente problema.

$$\text{Maximizar } Z = 5x_1 - x_1^2 + 8x_2 - x_2^2 + 10x_3 - x_3^2 + 15x_4 - x_4^2 + 20x_5 - x_5^2,$$

sujeta a

$x_1 \in \{3, 6, 12\}$, $x_2 \in \{3, 6\}$, $x_3 \in \{3, 6, 9, 12\}$,
$x_4 \in \{6, 12\}$, $x_5 \in \{9, 12, 15, 18\}$,
todas estas variables deben tener valores diferentes,
$x_1 + x_3 + x_4 \leq 25$.

Utilice las técnicas de programación de restricciones (reducción del dominio, propagación de restricciones, procedimiento de búsqueda y enumeración) para identificar todas las soluciones factibles y después encontrar una solución óptima. Muestre su trabajo.

11.9-3. Considere el siguiente problema.

$$\text{Maximizar } Z = 100x_1 - 3x_1^2 + 400x_2 - 5x_2^2 + 200x_3 - 4x_3^2 + 100x_4 - 2x_4^4,$$

sujeta a

$x_1 \in \{25, 30\}$, $x_2 \in \{20, 25, 30, 35, 40, 50\}$,
$x_3 \in \{20, 25, 30\}$, $x_4 \in \{20, 25\}$,
todas estas variables deben tener valores diferentes,
$x_2 + x_3 \leq 60$,
$x_1 + x_3 \leq 50$.

Utilice las técnicas de programación de restricciones (reducción del dominio, propagación de restricciones, procedimiento de búsqueda y enumeración) para identificar todas las soluciones factibles y después encontrar una solución óptima. Muestre su trabajo.

11.9-4. Considere el ejemplo de la Job Shop Co., que se presentó en la sección 8.3. En la tabla 8.25 se muestra su formulación como un problema de asignación. Use *restricciones globales* para formular un modelo de programación de restricciones compacto para este problema de asignación.

11.9-5. Considere el problema de asignar nadadores a las diferentes pruebas del equipo de natación que se presentó en el problema 8.3-4. En las respuestas al final del libro se muestra la formulación de este problema como uno de asignación. Utilice *restricciones globales* para formular un modelo de programación de restricciones compacto de este problema de asignación.

11.9-6. Considere el problema en el que se trata de determinar el mejor plan de horas dedicadas al estudio para cada uno de cuatro exámenes finales, que se presentó en el problema 10.3-3. Formule un modelo de programación de restricciones compacto de este problema.

11.9-7. El problema 10.3-2 señala que el propietario de una cadena de tres supermercados necesita determinar cuántas cargas de fresas frescas debe asignar a cada una de las tiendas. Formule un modelo de programación de restricciones compacto para representar este problema.

11.9-8. Una característica poderosa de la programación de restricciones es que las variables se pueden usar como subíndices de los términos de la función objetivo. Por ejemplo, considere el siguiente *problema del agente viajero*. El agente de ventas necesita visitar cada una de las n ciudades (ciudad $1, 2, \ldots, n$) sólo una vez, si comienza en

la ciudad 1 (su lugar de residencia) y regresa a la ciudad 1 después de completar el viaje. Sea c_{ij} la distancia desde la ciudad i hasta la ciudad j para $i, j = 1, 2, \ldots, n$ $(i \neq j)$. El objetivo es determinar cuál ruta debe seguir el vendedor para minimizar la distancia total del viaje. (Como se explica en el capítulo 13, este problema del agente viajero es un famoso problema clásico de IO con muchas aplicaciones que no tienen nada que ver con los agentes de ventas.)

Si la variable de decisión x_j $(j = 1, 2, \ldots, n, n + 1)$ denota la j-ésima ciudad visitada por el agente viajero, donde $x_1 = 1$ y $x_{n+1} = 1$, la programación de restricciones permite escribir el objetivo como

$$\text{Minimizar } Z = \sum_{j=1}^{n} c_{x_j x_{j+1}}.$$

Utilice esta función objetivo para formular un modelo completo de programación de restricciones para representar el problema.

11.10-1. De la parte inferior de las referencias seleccionadas que se proporcionaron al final del capítulo, seleccione una de las aplicaciones de la programación entera ganadora de premios. Lea este artículo y después escriba un resumen de dos páginas de la aplicación y los beneficios que brinda (incluya los no financieros).

11.10-2. De la parte inferior de las referencias seleccionadas que se proporcionaron al final del capítulo, seleccione tres de las aplicaciones de la programación entera ganadora de premios. Léalos y después escriba un resumen de una página de cada aplicación y los beneficios que brinda (incluya los no financieros).

CASOS

Caso 11.1 Aspectos de capacidad

Bentley Hamilton coloca la sección de negocios de *The New York Times* sobre la mesa de conferencias y observa cómo sus asociados se mueven incómodos en sus sillones.

El señor Hamilton desea que entiendan algo.

Pone la primera plana de *The Wall Street Journal* junto a la de *The New York Times* y observa a sus asociados abrir los ojos que ya no muestran aburrimiento.

El señor Hamilton quiere que entiendan bien.

Después arroja la primera página del *Financial Times* encima de los otros y observa que sus asociados comienzan a sudar.

El señor Hamilton quiere que lo que ocurre se grabe en las mentes de sus asociados.

"Les acabo de presentar tres periódicos financieros líderes con los chismes de negocios de hoy", declara con voz irritada. "Queridos asociados, ¡nuestra compañía se está yendo a pique! ¿Debo leerles los encabezados? De *The New York Times*: 'CommuniCorp baja al mínimo en 52 semanas'. De *The Wall Street Journal*: 'CommuniCorp pierde 25% del mercado de localizadores en sólo un año'. ¡Ah!, y en mi favorito, el *Financial Times*: 'CommuniCorp no puede ComuniCar: las acciones de CommuniCorp bajan debido a problemas de comunicación internos'. ¿Por qué la compañía cayó en una situación tan terrible?"

El señor Hamilton presenta una diapositiva que muestra una recta con una pendiente ligera hacia arriba. "Ésta es una gráfica de nuestra productividad en los últimos 12 meses. Como pueden ver, la productividad de la planta de localizadores ha aumentado en el último año. Es obvio que la productividad no es la causa del problema."

El señor Hamilton muestra una segunda diapositiva de una recta con una pendiente pronunciada hacia arriba. "Ésta es la gráfica de las órdenes que entregamos tarde o perdimos en los últimos 12 meses." Se escuchan movimientos incómodos de los asociados. "Como pueden ver en la gráfica, las órdenes tardías o perdidas han aumentado de manera firme y significativa en los últimos 12 meses. Creo que esta tendencia explica por qué ha caído nuestro porcentaje de mercado a su nivel más bajo en 52 semanas. Hemos hecho enojar y perdido a los distribuidores, clientes que dependían de entregas a tiempo para cumplir con las demandas de sus clientes."

"¿No cumplimos con las fechas de entrega cuando nuestro nivel de productividad debió permitir cumplir con las órdenes?", pregunta el señor Hamilton. "Hice esta pregunta a varios departamentos."

"¡Resulta que hemos producido localizadores por producirlos!", dice sin poderlo creer. "Los departamentos de mercadotecnia y ventas no se comunican con el de manufactura, por lo que los ejecutivos de éste último departamento no saben qué tipos de localizadores deben producir para cumplir con las órdenes. Ellos quieren que la planta siga trabajando, y producen localizadores sin importar si los han ordenado. Los localizadores terminados se envían al almacén, pero los ejecutivos de ventas y mercadotecnia no saben qué cantidades y estilos hay en él. Tratan de comunicarse con los ejecutivos del almacén para determinar si los localizadores inmovilizados en el inventario pueden satisfacer las órdenes, pero rara vez reciben una respuesta a sus preguntas."

El señor Hamilton hace una pausa y mira directo a sus asociados. "Señores y señoras, me parece que tenemos un problema de comunicación interna serio. Intento corregir este problema de inmediato. Quiero iniciar la instalación de una red de cómputo en toda la compañía para asegurar que todos los departamentos tengan acceso a documentos críticos y puedan comunicarse con facilidad entre sí por el correo electrónico. Como esta red interna (o intranet) representará un cambio importante en la infraestructura actual de comunicación, espero algunas fallas en el sistema y cierta resistencia de los empleados. Por todo ello, quiero que la instalación de la red se haga por etapas."

El señor Hamilton pasa las siguientes tablas de programación y requerimientos a sus asociados (donde IN = intranet).

Mes 1	Mes 2	Mes 3	Mes 4	Mes 5
Capacitación IN				
	Instalar IN en Ventas			
		Instalar IN en Manufactura		
			Instalar IN en Almacén	
				Instalar IN en Mercadotecnia

Departamento	Número de empleados
Ventas	60
Manufactura	200
Almacén	30
Comercialización	75

El señor Hamilton procede a explicar la tabla de tiempos y requerimientos. "No quiero que en el primer mes se hagan conexiones a la red; sólo deseo diseminar la información al respecto y tratar de que los empleados la acepten. El segundo mes se conectará el departamento de ventas, ya que recibe toda la información crítica de nuestros clientes. El tercer mes se conectará el departamento de manufactura. Durante el cuarto mes se instalará la red en el almacén y en el quinto y último quiero incluir al departamento de mercadotecnia en la red. La tabla de requerimientos que sigue a la programación contiene el número de empleados que requieren acceso a la red en cada departamento."

Después se dirige a Emily Jones, jefa de administración de información corporativa. "Necesito tu ayuda para planear la instalación de la red. En particular, la compañía necesita comprar servidores para la red interna. Los empleados se conectarán a los servidores y bajarán información a sus computadoras."

El señor Hamilton entrega a Emily la siguiente tabla con detalles de los tipos de servidores disponibles, el número de empleados que acepta cada servidor y los costos.

Tipo de servidor	Número de empleados que puede atender el servidor	Costo del servidor
PC con Intel Pentium estándar	Hasta 30 empleados	\$ 2 500
PC con Intel Pentium mejorado	Hasta 80 empleados	\$ 5 000
Estación de trabajo SGI	Hasta 200 empleados	\$10 000
Estación de trabajo Sun	Hasta 2 000 empleados	\$25 000

"Emily, necesito que decidas qué servidores comprar y cuándo comprarlos para minimizar el costo y asegurar que la compañía tiene suficiente capacidad de servidores para la implantación de la red interna a tiempo", dice el señor Hamilton. "Por ejemplo, puedes decidir comprar un servidor grande el primer mes para que acepte a todos los empleados o comprar un servidor pequeño cada mes de manera que un nuevo grupo de empleados tenga acceso a la red."

"Existen varios factores que complican tu decisión", continúa el señor Hamilton: "Dos fabricantes de servidores están dispuestos a ofrecer descuentos a CommuniCorp. SGI nos dará un descuento de 10% en cada servidor que compremos, pero sólo si se compra durante el primero o segundo meses. Sun está dispuesto a dar 25% de descuento en todos los servidores comprados los dos primeros meses. También tienes limitaciones en la cantidad de dinero que puedes gastar el primer mes. CommuniCorp ya ha asignado el presupuesto para los próximos dos meses, por lo cual sólo cuentas con un total de \$9 500 para las compras en los meses 1 y 2. Por último, el departamento de manufactura requiere al menos uno de los tres servidores más poderosos. Quiero tener tu decisión en mi escritorio para el fin de semana."

a) En primer lugar, Emily decide evaluar el número y tipo de servidores que va a comprar cada mes. Formule un modelo de PE para determinar qué servidores debe comprar Emily cada mes para minimizar los costos de ese periodo y dar el apoyo a los nuevos usuarios. ¿Cuántos servidores y de qué tipos debe comprar cada mes? ¿Cuál es el costo total del plan?

b) Emily se da cuenta de que tal vez logre algunos ahorros si compra un servidor más grande en los primeros meses para dar acceso a los usuarios en los últimos meses. Por lo tanto, decide evaluar el número y tipo de servidores que debe comprar durante todo el horizonte de planeación. Formule un modelo de PE para determinar qué servidores debe comprar Emily y en qué meses, de manera que se minimice el costo total y se dé acceso a todos los nuevos usuarios. ¿Cuántos servidores y de qué tipos debe comprar cada mes? ¿Cuál es el costo total del plan?

c) ¿Por qué es diferente la respuesta que se obtiene según el primer método que la que ofrece el segundo?

d) ¿Existen otros costos que Emily no ha tomado en cuenta en la formulación de su modelo? Si es así, ¿cuáles son?

e) ¿Qué otros aspectos pueden preocupar a los distintos departamentos de CommuniCorp respecto de la red interna?

RESUMEN DE LOS CASOS ADICIONALES EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

CASO 11.2 Asignación de arte

Se realizan planes para llevar a cabo una exposición de artistas modernos emergentes en el museo de arte moderno de San Francisco. Se ha recopilado una larga lista de posibles artistas, sus obras disponibles y los precios nominales de estas obras. También existen varias restricciones en relación con la mezcla de obras que se pueden elegir. El siguiente paso es aplicar PEB para hacer la selección de las obras para la exposición bajo tres escenarios diferentes.

CASO 11.3 Juegos de cocina en almacén

Una mala administración de inventarios en el almacén local de Furniture City ha causado la acumulación excesiva de muchos artículos y el déficit de algunos otros. Para comenzar a rectificar esta situación, se han identificado los 20 juegos de cocina más populares en el departamento correspondiente. Estos juegos se componen de hasta ocho características en diversos estilos, de forma que cada uno de dichos estilos debe tener un correcto nivel de inventario en el almacén. Sin embargo, la cantidad limitada de espacio de almacenaje asignado al departamento de cocina significa que se deben tomar algunas decisiones difíciles en este aspecto. Después de recopilar los datos relevantes de los 20 juegos de cocina, ahora es necesario aplicar PEB para determinar qué cantidad de características y estilos de Furniture City se deben conservar en el almacen local bajo tres escenarios diferentes.

CASO 11.4 Asignación de estudiantes a escuelas (de nuevo)

Como se presentó en el caso 4.3 y se revisó en el caso 6.3, la oficina escolar de Springfield necesita asignar a los estudiantes de educación media de las seis áreas residenciales de la ciudad en las tres escuelas que aún se conservan. La nueva complicación de la oficina escolar es que ha decidido no permitir que los estudiantes de una misma área se dividan entre distintas escuelas. Por lo tanto, como cada una de las seis áreas debe ser asignada a una sola escuela, ahora se debe aplicar PEB para hacer estas asignaciones bajos los distintos escenarios considerados en el caso 4.3.

CAPÍTULO 12

Programación no lineal

El papel fundamental de la programación lineal en IO se refleja con exactitud en el hecho de que es el tema central de *un tercio* de este libro. Un supuesto importante de programación lineal es que *todas sus funciones* (objetivo y de restricción) son lineales. Aunque, en esencia, este supuesto se cumple en el caso de muchos problemas prácticos, con frecuencia no es así. Por lo tanto, muchas veces es necesario manejar problemas de programación no lineal, que es el área importante en la que se enfocará la atención en este capítulo.

De manera general,[1] el *problema de programación no lineal* consiste en encontrar $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ para

$$\text{maximizar} \qquad f(\mathbf{x}),$$

sujeta a

$$g_i(\mathbf{x}) \leq b_i, \qquad \text{para } i = 1, 2, \ldots, m,$$

y

$$\mathbf{x} \geq \mathbf{0},$$

donde $f(\mathbf{x})$ y $g_i(\mathbf{x})$ son funciones dadas de n variables de decisión.[2]

Existen muchos tipos de problemas de programación no lineal, lo cual depende de las características de las funciones $f(\mathbf{x})$ y $g_i(\mathbf{x})$. Se emplean varios algoritmos para resolver los diferentes tipos de problemas. Para ciertos tipos donde las funciones tienen formas simples, los problemas pueden resolverse de manera relativamente eficiente. En algunos otros casos, incluso la solución de pequeños problemas representa un verdadero reto.

Debido a que existen muchos tipos de problemas y muchos algoritmos, la programación no lineal es un tema muy extenso. No se cuenta con el espacio suficiente para realizar un estudio completo. Sin embargo, se presentarán algunos ejemplos de aplicación y después se introducirán algunas ideas básicas para resolver ciertos tipos importantes de problemas de programación no lineal.

En los apéndices 2 y 3 se proporcionan antecedentes útiles para este capítulo y se recomienda consultarlos al estudiar las secciones que siguen.

12.1 APLICACIONES DE MUESTRA

Los siguientes ejemplos ilustran unos cuantos de los muchos tipos importantes de problemas a los que se aplica la programación no lineal.

[1] Las otras *formas legítimas* corresponden a aquellas para *programación lineal* que se mencionan en la sección 3.2. La sección 4.6 describe cómo convertir estas otras formas en la forma que se proporciona aquí.

[2] Para simplificar, se supondrá a lo largo del capítulo que *todas* estas funciones o bien son *diferenciables* en todas partes o son funciones lineales por partes (tema que se estudió en las secciones 12.1 y 12.8).

Problema de mezcla de productos con elasticidad de precios

En problemas de *mezcla de productos*, como el de la Wyndor Glass Co., de la sección 3.1, la meta es determinar la mezcla óptima de los niveles de elaboración de los productos de una empresa, dadas las limitaciones sobre los recursos que se necesitan para manufacturarlos, con el objeto de maximizar la ganancia total de la empresa. En algunos casos existe una ganancia unitaria fija asociada a cada producto, con lo cual la función objetivo que se obtiene es lineal. Sin embargo, en muchos problemas de mezcla de productos, ciertos factores introducen no linealidades en la función objetivo.

Por ejemplo, un gran fabricante puede encontrar *elasticidad de precios*, es decir, que la cantidad de un producto que se puede vender tiene relación inversa con el precio que se cobra por él. Así, la *curva precio-demanda* de un producto puede parecerse a la que se muestra en la figura 12.1, donde $p(x)$ es el precio que se necesita para poder vender x unidades. Como sabemos, la ganancia de la empresa por producir y vender x unidades es el ingreso que generan las ventas $xp(x)$ menos los costos de elaboración y distribución. Por lo tanto, si el costo unitario de producir y distribuir el artículo está fijo en c (vea la línea punteada de la figura 12.1), la ganancia de la empresa por producir y vender x unidades está dada por la función no lineal

$$P(x) = xp(x) - cx,$$

como se puede observar en la figura 12.2. Si *cada uno* de los n productos de la empresa tiene una función de ganancia parecida, por ejemplo $P_j(x_j)$ por producir y vender x_j unidades del producto j ($j = 1, 2, \ldots, n$), entonces la función objetivo global es

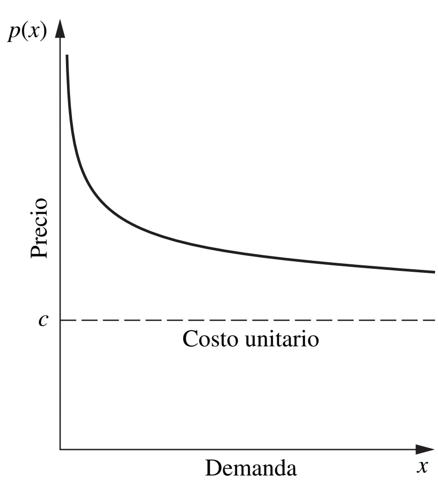
$$f(\mathbf{x}) = \sum_{j=1}^{n} P_j(x_j),$$

esto es, una suma de funciones no lineales.

Otra razón por la que pueden surgir no linealidades en la función objetivo es que el *costo marginal* de producir otra unidad más de un producto dado puede variar según el nivel de elaboración. Por ejemplo, el costo marginal puede disminuir cuando aumenta el nivel de producción, gracias al *efecto de curva de aprendizaje* (mayor eficiencia a medida que aumenta la experiencia). Por otro lado, pueden aumentar a causa de ciertas medidas especiales, como tiempos extra o instalaciones fabriles más costosas, necesarias para aumentar la producción.

La no linealidad también puede surgir en las funciones de restricción $g_i(\mathbf{x})$ en forma bastante parecida. Por ejemplo, si existe una restricción de presupuesto sobre el costo total de producción, la función de costo será no lineal si el costo marginal de producción varía como se acaba de describir. En el caso de restricciones sobre otro tipo de recursos, $g_i(\mathbf{x})$ será no lineal siempre que el uso del recurso correspondiente no sea estrictamente proporcional a los niveles de los respectivos productos.

FIGURA 12.1
Curva precio-demanda.



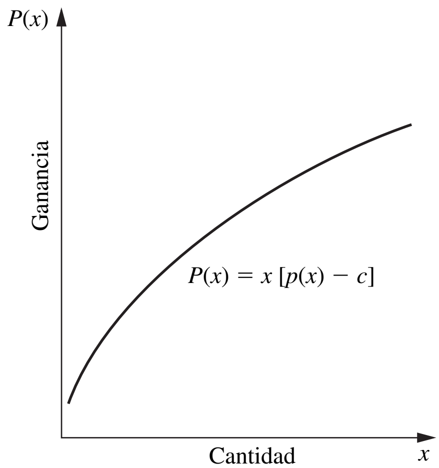
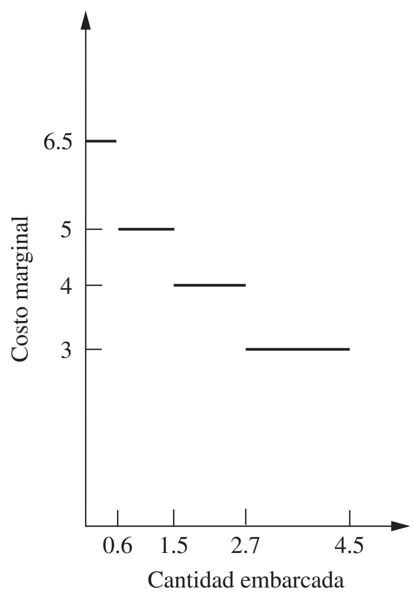


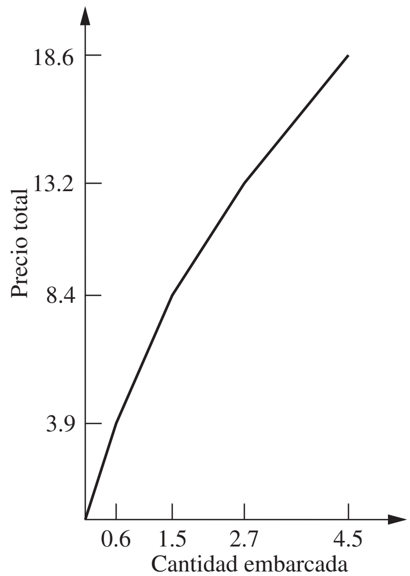
Figura 12.2
Función de ganancia.

Problema de transporte con descuento de precios por volumen de embarque

Como se ilustró en el ejemplo de la P & T Company, en la sección 8.1, una aplicación común del problema de transporte es determinar un plan óptimo para enviar bienes desde varios orígenes hasta varios destinos, dadas las restricciones de recursos y demanda, con el fin de minimizar el costo total de transporte. En el capítulo 8 se supuso que el *costo por unidad enviada* de un origen a un destino dados *es fijo*, independientemente de la cantidad que se envíe. En la realidad, este costo puede no ser fijo. A veces se dispone de *descuentos por cantidad* para volúmenes grandes, con lo que el *costo marginal* de enviar una unidad más puede seguir un patrón como el que se muestra en la figura 12.3. En consecuencia, el costo que resulta de embarcar x unidades está dado por la función *no lineal* $C(x)$ que es una *función lineal por partes* con pendiente igual al costo marginal, como la que se muestra en la figura 12.4. [La función de la figura 12.4 consiste en un segmento de recta con pendiente 6.5 desde (0, 0) hasta (0.6, 3.9), un segundo segmento de recta con pendiente 5 desde (0.6, 3.9) hasta (1.5, 8.4), un tercer segmento de recta con pendiente 4 desde (1.5, 8.4) hasta (2.7, 13.2) y uno más con pendiente 3 de (2.7, 13.2) hasta (4.5, 18.6)]. En consecuencia, si *cada*

FIGURA 12.3
Costo marginal de transporte.





■ **FIGURA 12.4**
Función del costo de transporte.

combinación de origen y destino tiene una función de costos similar, es decir, si el costo de enviar x_{ij} unidades del origen $i(i = 1, 2, \ldots, m)$ al destino j $(j = 1, 2, \ldots, n)$ está dado por una función no lineal $C_{ij}(x_{ij})$, entonces la función objetivo global que se va a minimizar es

$$f(\mathbf{x}) = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij}(x_{ij}).$$

Aun con esta función objetivo no lineal, es normal que las restricciones sean del tipo lineales especiales que se ajustan al modelo del problema de transporte de la sección 8.1.

Selección de una cartera de inversiones riesgosas

En la actualidad, una práctica común entre los administradores de grandes carteras de inversión es usar, como guía, modelos de computadora basados en parte en programación no lineal. Debido a que los inversionistas se preocupan tanto por el *rendimiento esperado* (ganancia), como por el *riesgo* asociado con su inversión, la programación no lineal se emplea para determinar una cartera que, bajo ciertos supuestos, proporcione un trueque óptimo entre estos dos factores. Este enfoque se basa en gran parte en la novedosa investigación realizada por Harry Markowitz y William Sharpe, que los ayudó a ganar el Premio Nobel de Economía en 1990.

Se puede formular un modelo de programación no lineal para representar este problema de la siguiente manera. Suponga que se están tomando en cuenta n tipos de acciones para incluirlas en la cartera; las variables de decisión $x_j (j = 1, 2, \ldots, n)$ representan el número de acciones j que se van a incluir. Sean μ_j y σ_{jj} la *media* y la *varianza* (estimadas) del rendimiento de cada acción del tipo j, donde σ_{jj} mide el riesgo de estas acciones. Para $i = 1, 2, \ldots, n$ $(i \neq j)$, sea σ_{ij} la *covarianza* del rendimiento de una acción de cada tipo i y tipo j. (Como sería difícil estimar todas las σ_{ij}, lo usual es basarse en ciertos supuestos sobre el comportamiento del mercado que permitan calcular σ_{ij} directamente, a partir de σ_{ii} y σ_{jj}.) Entonces, el valor esperado de $R(\mathbf{x})$ y la varianza $V(\mathbf{x})$ del rendimiento total de la cartera completa son

$$R(\mathbf{x}) = \sum_{j=1}^{n} \mu_j x_j$$

y

$$V(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j,$$

Recuadro de aplicación

Bank Hapoalim Group es la corporación bancaria más grande de Israel, y brinda servicios en ese país a través de una red de 327 sucursales, nueve centros de negocios regionales y varias subsidiarias nacionales. También realiza operaciones a nivel mundial a través de 37 sucursales, oficinas y subsidiarias en los centros financieros más importantes de Norte y Sudamérica, así como en Europa.

Una parte fundamental del negocio de Bank Hapoalim radica en brindar asesoría en inversiones a sus clientes. Para conservar ventaja sobre sus competidores, la cúpula administrativa se ha dado a la tarea de realizar un programa de reestructuración para ofrecer a dichos asesores en inversiones la tecnología y metodología más moderna. Por esta razón se integró un grupo de investigación de operaciones.

Este grupo llegó a la conclusión de que necesitaba desarrollar un sistema de soporte a las decisiones que fuera flexible para los asesores en inversiones que pudiera ser adaptable para satisfacer las diferentes necesidades de cada cliente. A cada uno de estos clientes se les solicitó proporcionar una gran cantidad de información acerca de sus necesidades, lo que incluía la elección entre las diferentes alternativas con respecto a sus objetivos de inversión, horizonte de inversión, selección de un índice para esforzarse a exceder la meta, preferencia con respecto a la liquidez y tipo de cambio, etc. También se les planteó una serie de preguntas con el fin de conocer la clasificación del nivel de riesgo que se atrevían a asumir.

La elección natural del modelo para manejar el sistema de soporte a las decisiones resultante (llamado *Opti-Money System*) fue el *modelo de programación no lineal clásico para la selección del portafolio* descrito en esta sección del libro, con algunas modificaciones que incorporan toda la información acerca de las necesidades de cada uno de los clientes. Este modelo genera una ponderación óptima de 60 posibles clases de activos de fondos y bonos para el portafolio y el asesor de inversiones, en conjunto con el cliente, selecciona los fondos y bonos específicos dentro de esas clases.

En un año reciente, los asesores en inversiones del banco, participaron en 133 000 sesiones de consulta con 63 000 clientes, utilizando este sistema de soporte a las decisiones. Las *ganancias anuales* con respecto a los puntos de referencia de los clientes que siguieron el consejo del sistema sumaron un *total de alrededor de* **244 millones de dólares**, *a la vez que aportaron más de* **31 millones de dólares** *al ingreso anual del banco*.

Fuente: M. Avriel, H. Pri-Zan, R. Meiri y A. Peretz, "Opti-Money at Bank Hapoalim: A Model-Based Investment Decision-Support System for Individual Customers", en *Interfaces*, **34**(1): 39-50, enero-febrero de 2004. (En la página en internet de este libro www.mhhe.com/hillier se proporciona una liga hacia este artículo.)

donde $V(\mathbf{x})$ mide el riesgo asociado con la cartera. Una manera de considerar el trueque entre estos dos factores es usar $V(\mathbf{x})$ como la función objetivo que debe minimizarse y después imponer la restricción de que $R(\mathbf{x})$ no debe ser menor que el rendimiento mínimo aceptable esperado. El modelo de programación no lineal completo sería

$$\text{Minimizar} \qquad V(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j,$$

sujeta a

$$\sum_{j=1}^{n} \mu_j x_j \geq L$$

$$\sum_{j=1}^{n} P_j x_j \leq B$$

y

$$x_j \geq 0, \qquad \text{para } j = 1, 2, \ldots, n,$$

donde L es el rendimiento mínimo aceptable esperado, P_j es el precio de cada acción tipo j y B es la cantidad de dinero presupuestada para la cartera.

Uno de los defectos de esta formulación es que existe una dificultad relativa al elegir un valor adecuado de L para obtener la mejor combinación entre $R(\mathbf{x})$ y $V(\mathbf{x})$. Entonces, en lugar de terminar el proceso con una elección de L, es común usar un enfoque de programación (no lineal) *paramétrica* para generar la solución óptima como una función de L en un intervalo amplio de valores; el siguiente paso consiste en analizar los valores de $R(\mathbf{x})$ y $V(\mathbf{x})$ de estas soluciones que son óptimas para algún valor de L y después elegir la solución que dé la mejor opción para estas dos cantidades. Con frecuencia se hace referencia a este procedimiento como la generación de soluciones sobre la *frontera eficiente* de la gráfica de dos dimensiones de los puntos $R(\mathbf{x})$, $V(\mathbf{x})$ para todas las $\mathbf{x}$ factibles.

La razón es que el punto $R(\mathbf{x})$, $V(\mathbf{x})$ para una **x** óptima (para alguna L) se encuentra sobre la *frontera* (límite) de los puntos factibles. Más aún, cada punto **x** es *eficiente* en el sentido de que ninguna otra solución factible será cuando menos tan buena con una medida (R o V) y estrictamente mejor con la otra medida (V más pequeña o R más grande).

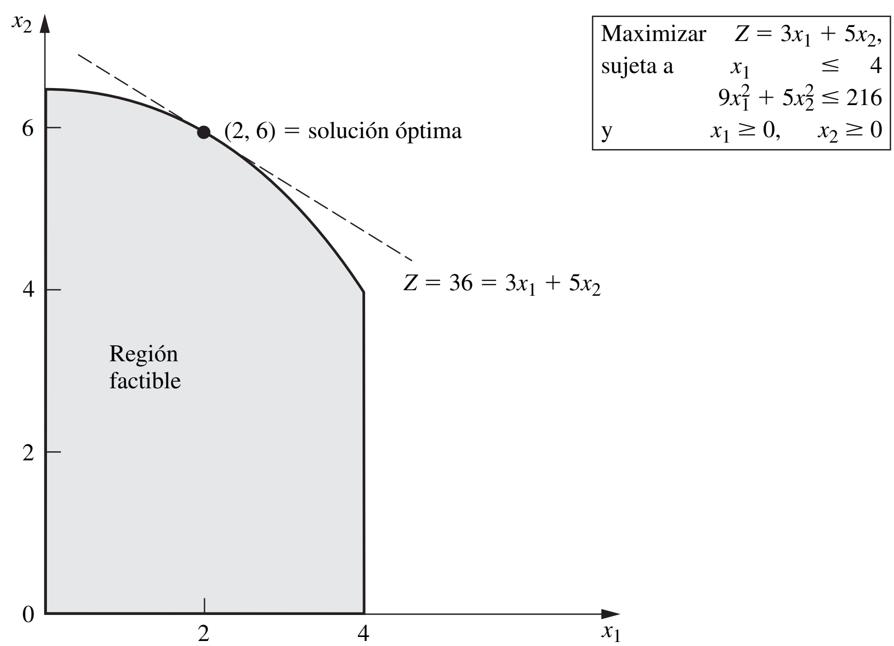
Esta aplicación de la programación no lineal es particularmente importante. En la actualidad, el uso de la programación no lineal para optimizar el portafolio forma una parte medular del análisis financiero moderno. (En palabras más generales, el relativamente novedoso campo de la ingeniería financiera ha surgido con el fin de otorgar la importancia que le corresponde a la aplicación de las técnicas de investigación de operaciones, tales como la programación no lineal, a diferentes problemas financieros, dentro de los que se incluye la optimización del portafolio.) Como lo ilustra este recuadro de aplicación, este tipo de uso de la programación no lineal ha provocado un sorprendente efecto en la práctica. De manera continua se lleva a cabo un gran esfuerzo en el área de investigación para ampliar el conocimiento acerca de las propiedades y las aplicaciones del modelo anterior y los modelos de programación no lineal relacionados en varios tipos de complejos análisis de portafolio.[3]

12.2 ILUSTRACIÓN GRÁFICA DE PROBLEMAS DE PROGRAMACIÓN NO LINEAL

Cuando un problema de programación no lineal tiene sólo una o dos variables, se puede representar en forma gráfica de una manera muy parecida al ejemplo de la Wyndor Glass Co., de programación lineal, de la sección 3.1. Se verán unos cuantos ejemplos, puesto que una representación gráfica de este tipo proporciona una visión global de las propiedades de las soluciones óptimas de programación lineal y no lineal. Con el fin de hacer hincapié en las diferencias entre programación lineal y no lineal, se usarán algunas variaciones no lineales del problema de la Wyndor Glass Co.

En la figura 12.5 se muestra lo que ocurre con este problema si los únicos cambios que se hacen al modelo de la sección 3.1 son que la segunda y tercera restricciones funcionales se sustituyen por la restricción no lineal $9x_1^2 + 5x_2^2 \leq 216$. Compare las figuras 12.5 y 3.3. La solución óptima sigue

FIGURA 12.5
Ejemplo de la Wyndor Glass Co., con la restricción no lineal $9x_1^2+5x_2^2 \leq 216$ en lugar de la segunda y tercera restricción original.



[3] A continuación se mencionan investigaciones recientes en este campo. B. I. Jacobs, K. N. Levy y H. M. Markowitz, "Portfolio Optimization with Factors, Scenarios and Realistic Short Positions", en *Operations Research*, **53**(4): 586-599, julio-agosto de 2005: A. F. Siegel y A. Woodgate, "Performance of Portfolios Optimized with Estimation Error", en *Management Science*, **53**(6): 1005-1015, junio de 2007; H. Konno y T. Koshizuka, "Mean-Absolute Deviation Model", en *IEE Transactions*, **37**(10): 893-900, octubre de 2005.

siendo $(x_1, x_2) = (2, 6)$. Aún más, todavía se encuentra sobre la frontera de la región factible, pero *no* es una solución factible en un vértice (FEV). La solución óptima pudo haber sido FEV con una función objetivo diferente (verifique $Z = 3x_1 + x_2$), pero el hecho de que no necesite serlo significa que ya no se puede aprovechar la gran simplificación que se utiliza en programación lineal que permite limitar la búsqueda de una solución óptima para las soluciones FEV.

Ahora suponga que las restricciones lineales de la sección 3.1 se conservan sin cambio, pero que la función objetivo se hace no lineal. Por ejemplo, si

$$Z = 126x_1 - 9x_1^2 + 182x_2 - 13x_2^2,$$

la representación gráfica de la figura 12.6 indica que la solución óptima es $x_1 = \frac{8}{3}$, $x_2 = 5$, que de nuevo se encuentra en la frontera de la región factible. (El valor óptimo de Z es $Z = 857$; así, la figura 12.6 muestra el hecho de que el lugar geométrico de todos los puntos con $Z = 857$ tiene en común con la región factible sólo este punto, mientras que el lugar geométrico de los puntos con Z más grande no toca la región factible en ningún punto.) Por otro lado, si

$$Z = 54x_1 - 9x_1^2 + 78x_2 - 13x_2^2,$$

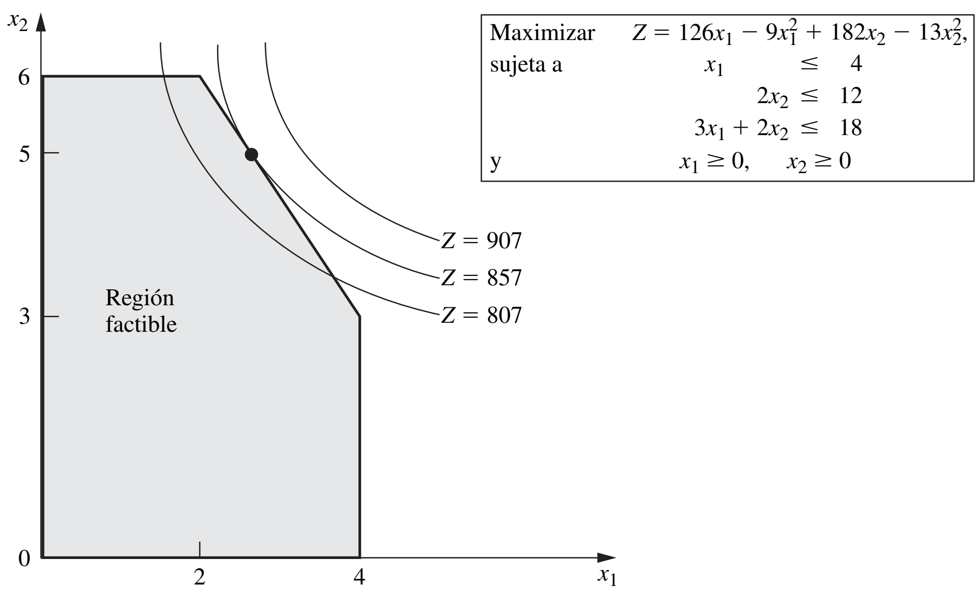
entonces la figura 12.7 ilustra que la solución óptima es $(x_1, x_2) = (3, 3)$, que se encuentra *dentro* de la frontera de la región factible. (Se puede comprobar que esta solución es óptima si se usa cálculo para derivarla como un máximo global no restringido; como también satisface las restricciones, debe ser óptima para el problema restringido.) Por lo tanto, es necesario que un algoritmo general para resolver problemas de este tipo tome en cuenta *todas* las soluciones en la región factible, y no sólo aquellas que están sobre la frontera.

Otra complicación que surge en programación no lineal es que un máximo *local* no necesariamente es un máximo *global* (la solución óptima global). Por ejemplo, considere la función de una sola variable graficada en la figura 12.8. En el intervalo $0 \leq x \leq 5$, esta función tiene tres máximos locales —$x = 0$, $x = 2$ y $x = 4$— pero sólo uno de éstos —$x = 4$— es un *máximo global*. (De igual manera, existen mínimos locales en $x = 1$, 3 y 5, pero sólo $x = 5$ es un *mínimo global*.)

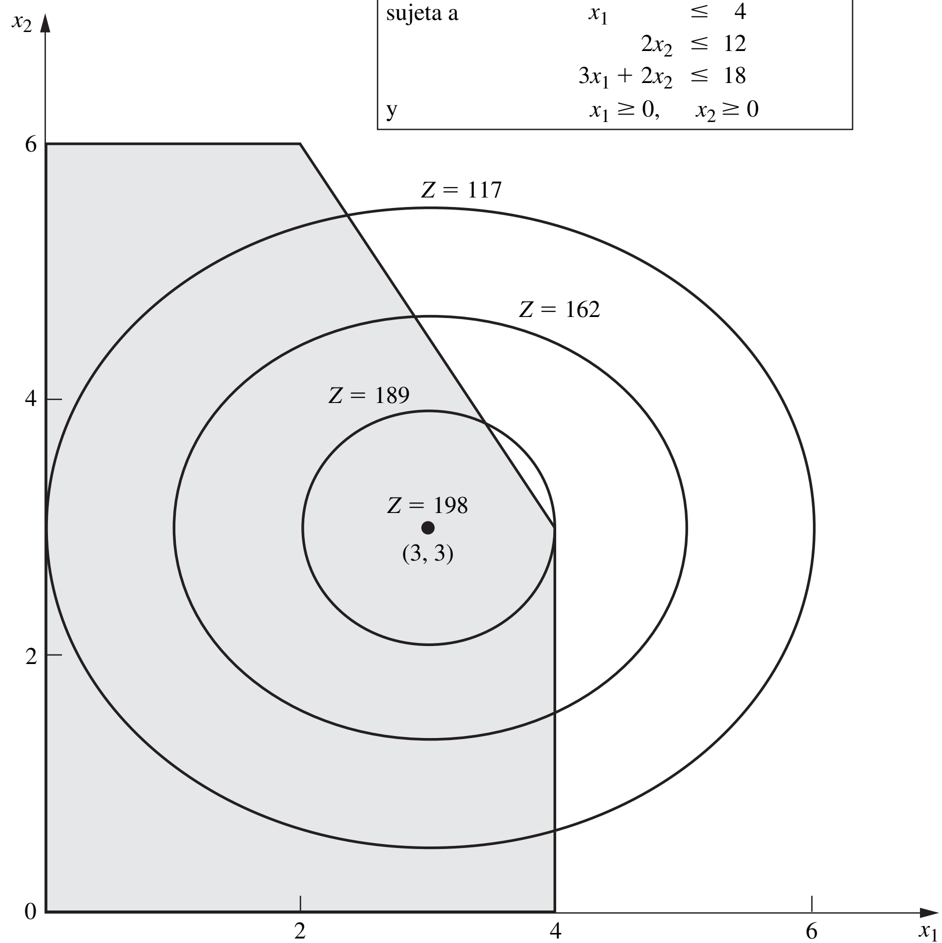
En general, los algoritmos de programación no lineal no pueden distinguir entre un máximo local y un máximo global (excepto si encuentran otro máximo local *mejor*). Por tanto, es crucial conocer las condiciones en las que se *garantiza* que un máximo local es un máximo global en la región factible. Recuerde que en cálculo, cuando se maximiza una función ordinaria (doblemente diferenciable) de una sola variable $f(x)$ sin restricciones, esta garantía está dada cuando

$$\frac{\partial^2 f}{\partial x^2} \leq 0 \qquad \text{para toda } x.$$

■ **FIGURA 12.6**
Ejemplo de la Wyndor Glass Co., con la región factible original pero con la función objetivo no lineal $Z = 126x_1 - 9x_1^2 + 182x_2 - 13x_2^2$ en lugar de la función objetivo original.



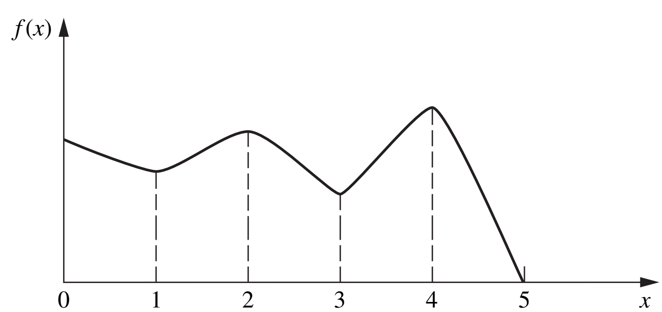
$$
\begin{array}{lrl}
\text{Maximizar} & Z = 54x_1 - 9x_1^2 + 78x_2 - 13x_2^2, & \\
\text{sujeta a} & x_1 & \leq 4 \\
& 2x_2 & \leq 12 \\
& 3x_1 + 2x_2 & \leq 18 \\
\text{y} & x_1 \geq 0, \quad x_2 \geq 0 &
\end{array}
$$



■ **FIGURA 12.7** Ejemplo de la Wyndor Glass Co., con la región factible original pero con otra función objetivo no lineal, $Z = 54x_1 - 9x_1^2 + 78x_2 - 13x_2^2$, en lugar de la original.

Una función de este tipo que siempre es "cóncava hacia abajo" (o que no tiene convexidad) se llama función **cóncava**.[4] De igual manera, si se sustituye $\leq$ por $\geq$, de manera que la función sea siempre "cóncava hacia arriba" (o que no tenga concavidad), se llama función **convexa**.[5] (Por ello, una función *lineal* puede ser tanto cóncava como convexa.) En la figura 12.9 se pueden ver este tipo de ejemplos. Observe que la figura 12.8 ilustra una función que no es cóncava ni convexa puesto que alterna sus curvaturas hacia arriba y hacia abajo.

■ **FIGURA 12.8** Una función con varios máximos locales ($x = 0, 2, 4$), pero sólo $x = 4$ es un máximo global.



[4] Las funciones *cóncavas* a veces se llaman *cóncavas hacia abajo*.

[5] Las funciones *convexas* a veces se llaman *cóncavas hacia arriba*.

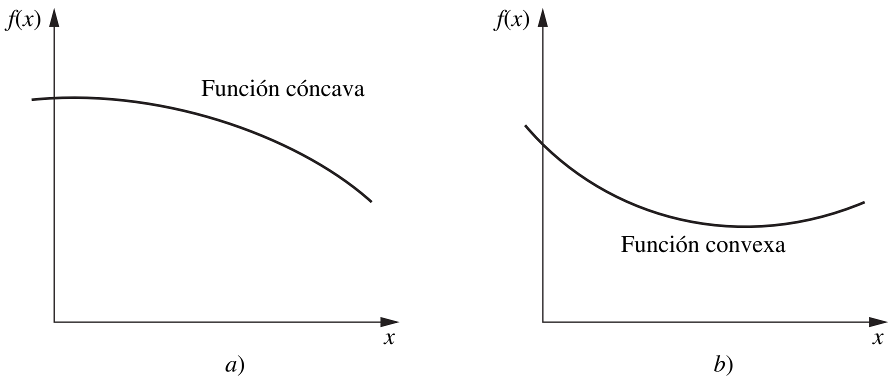


FIGURA 12.9
Ejemplos de *a*) una función cóncava y *b*) una función convexa.

Las funciones de variables múltiples también se pueden caracterizar como cóncavas o convexas si su curvatura es siempre hacia abajo o hacia arriba. Estas definiciones intuitivas se fundamentan en términos precisos que, junto con cierta profundización en los conceptos, se presentan en el apéndice 2. (Las funciones cóncava y convexa tienen un papel fundamental en la programación no lineal, por lo que, si no se está muy familiarizado con estas funciones, se recomienda leer a conciencia el apéndice 2.) Este apéndice también proporciona una prueba conveniente para verificar si una función de dos variables es cóncava, convexa o ninguna de las dos.

La siguiente es una forma práctica de verificar esta característica en el caso de una función de más de dos variables cuando la función consiste en una suma de funciones más pequeñas cada una de sólo una o dos variables. Si cada función más pequeña es cóncava, entonces la función completa es cóncava. De manera similar, la función completa es convexa si cada función más pequeña es convexa.

Para ilustrar esta afirmación considere la función

$$\begin{aligned} f(x_1, x_2, x_3) &= 4x_1 - x_1^2 - (x_2 - x_3)^2 \\ &= [4x_1 - x_1^2] + [-(x_2 - x_3)^2], \end{aligned}$$

que es la suma de las dos funciones más pequeñas dadas en los corchetes. La primera función más pequeña $4x_1 - x_1^2$ es una función sólo de la variable x_1, por lo que puede verse que es cóncava si se observa que su segunda derivada es negativa. La segunda función más pequeña $-(x_2 - x_3)^2$ es una función de x_2 y x_3, por lo que se puede aplicar la prueba para funciones de dos variables dada en el apéndice 2. En realidad, este apéndice usa esta función en particular para ilustrar la prueba y encuentra que la función es cóncava. Como las dos funciones más pequeñas son cóncavas, la función completa $f(x_1, x_2, x_3)$ también debe serlo.

Si un problema de programación no lineal no tiene restricciones, el hecho de que la función objetivo sea *cóncava* garantiza que un máximo local es un *máximo global*. (De igual manera, una función objetivo *convexa* asegura que un mínimo local es un *mínimo global*.) Si existen restricciones, se necesita una condición más para dar esta garantía, a saber, que la *región factible* sea un *conjunto convexo*. Por esta razón, los conjuntos convexos tienen un papel fundamental en la programación no lineal.

Como se analiza en el apéndice 2, un **conjunto convexo** es simplemente un conjunto de puntos tales que, para cada par de puntos de la colección, el segmento de recta que los une está totalmente contenido en ella. Así, la región factible en el problema original de la Wyndor Glass Co. (vea la figura 12.6 o 12.7) es un conjunto convexo. En realidad, la región factible de *cualquier* otro problema de programación lineal es un conjunto convexo. De igual manera, la región factible de la figura 12.5 también es un conjunto convexo.

En general, la región factible de un problema de programación no lineal es un conjunto convexo siempre que todas las funciones $g_i(\mathbf{x})$ [para las restricciones $g_i(\mathbf{x}) \le b_i$] sean convexas. En el ejemplo de la figura 12.5, las dos $g_i(\mathbf{x})$ son convexas, puesto que $g_1(\mathbf{x}) = x_1$ (una función lineal es automáticamente cóncava y convexa) y $g_2(\mathbf{x}) = 9x_1^2 + 5x_2^2$ (tanto $9x_1^2$ como $5x_2^2$ son funciones con-

vexas, por lo que su suma es una función convexa). Estas dos funciones convexas $g_i(\mathbf{x})$ conducen a que la región factible de la figura 12.5 sea un conjunto convexo.

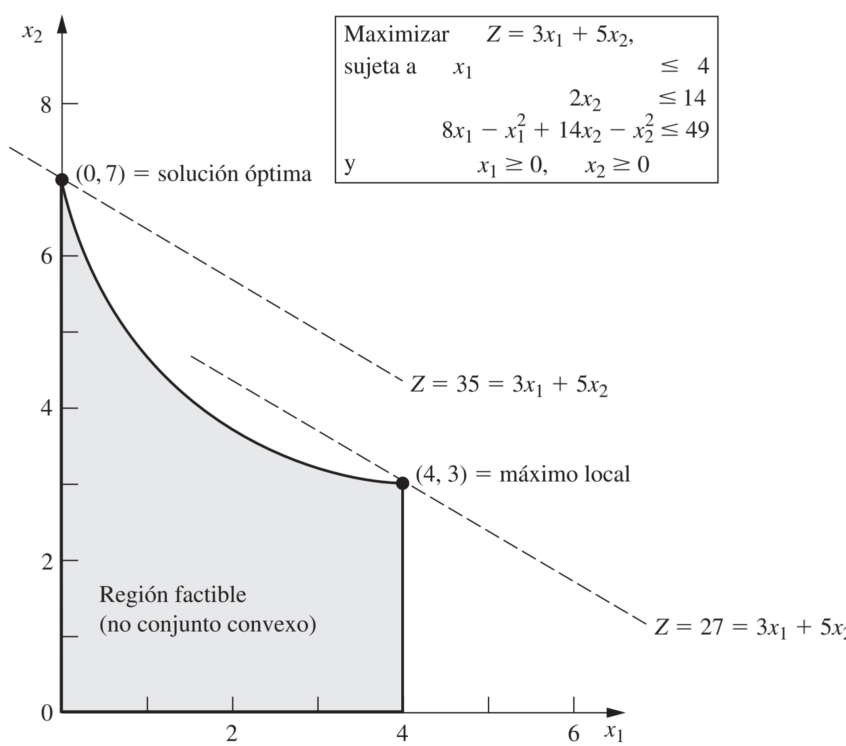
Ahora permítanos mostrarle qué sucede cuando sólo una de estas funciones $g_i(\mathbf{x})$ es una función cóncava. En particular, suponga que los únicos cambios que se introducen al ejemplo original de la Wyndor Glass Co., son que sus restricciones funcionales segunda y tercera se sustituyen por $2x_2 \leq 14$ y $8x_1 - x_1^2 + 14x_2 - x_2^2 \leq 49$. Por lo tanto, la nueva $g_3(\mathbf{x}) = 8x_1 - x_1^2 + 14x_2 - x_2^2$, que es una función cóncava ya que tanto $8x_1 - x_1^2$ como $14x_2 - x_2^2$ son funciones cóncavas. La nueva región factible que se muestra en la figura 12.10 *no* es un conjunto convexo. ¿Por qué? Porque contiene pares de puntos, como (0, 7) y (4, 3), tales que parte del segmento de recta que los une no está en la región factible. En consecuencia, no se puede garantizar que un máximo local sea un máximo global. En realidad, este ejemplo tiene dos máximos locales (0, 7) y (4, 3), pero sólo (0, 7) es un máximo global.

En consecuencia, para garantizar que un máximo local sea un máximo global en un problema de programación no lineal con restricciones $g_i(\mathbf{x}) \leq b_i$ $(i = 1, 2, \ldots, m)$ y $\mathbf{x} \geq \mathbf{0}$, la función objetivo $f(\mathbf{x})$ debe ser *cóncava* y cada $g_i(\mathbf{x})$ debe ser *convexa*. Un problema de este tipo se llama *problema de programación convexa* y es una de las clases más importantes de la programación no lineal que se estudiará en la sección 12.3.

12.3 TIPOS DE PROBLEMAS DE PROGRAMACIÓN NO LINEAL

Los problemas de programación no lineal se presentan de muchas formas distintas. Al contrario del método símplex para programación lineal, no se dispone de un algoritmo que resuelva todos estos tipos especiales de problemas. En su lugar, se han desarrollado algoritmos para algunas *clases* (tipos especiales) de problemas de programación no lineal. En esta sección se introducirán brevemente las clases más importantes. Después, en las secciones subsecuentes, se describirá cómo se pueden resolver algunos de estos problemas. Para simplificar la explicación, se supondrá a partir de ahora que los problemas se han formulado (o reformulado) en la forma general que se presentó al principio del capítulo.

FIGURA 12.10
Ejemplo de la Wyndor Glass Co., donde $2x_2 \leq 14$ y una restricción no lineal, $8x_1 - x_1^2 + 14x_2 - x_2^2 \leq 49$, reemplazan la segunda y tercera restricciones funcionales originales.



Optimización no restringida

Los problemas de optimización no restringida *no* tienen restricciones, por lo que la función objetivo es, sencillamente,

$$\text{Maximizar} \qquad f(\mathbf{x})$$

sobre *todos* los valores de $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Según el repaso del apéndice 3, la condición *necesaria* para que una solución específica $\mathbf{x} = \mathbf{x}^*$ sea óptima cuando $f(\mathbf{x})$ es una función diferenciable es

$$\frac{\partial f}{\partial x_j} = 0 \qquad \text{en } \mathbf{x} = \mathbf{x}^*, \text{ para } j = 1, 2, \ldots, n.$$

Cuando $f(\mathbf{x})$ es *cóncava*, esta condición también es *suficiente*, por lo que la obtención de $\mathbf{x}^*$ se reduce a resolver el sistema de las n ecuaciones que se obtuvieron al establecer las n derivadas parciales iguales a cero. Desafortunadamente, cuando se trata de funciones *no lineales* $f(\mathbf{x})$, estas ecuaciones también suelen ser *no lineales*, en cuyo caso es poco probable que se pueda obtener una solución analítica simultánea. ¿Qué se puede hacer en ese caso? En las secciones 12.4 y 12.5 se describen *procedimientos algorítmicos de búsqueda* para encontrar $\mathbf{x}^*$, primero para $n = 1$ y luego para $n > 1$. Estos procedimientos también tienen un papel importante en la solución de varios tipos de problemas con restricciones, que se describirán en seguida. La razón es que muchos algoritmos para problemas *restringidos* están construidos de forma que se adaptan a versiones *no restringidas* del problema en una parte de cada iteración.

Cuando una variable x_j tiene una restricción de no negatividad, $x_j \geq 0$, la condición necesaria y —tal vez— suficiente anterior cambia ligeramente a

$$\frac{\partial f}{\partial x_j} \begin{cases} \leq 0 & \text{para } \mathbf{x} = \mathbf{x}^*, \quad \text{si } x_j^* = 0 \\ = 0 & \text{para } \mathbf{x} = \mathbf{x}^*, \quad \text{si } x_j^* > 0 \end{cases}$$

para cada j de este tipo. Esta condición se ilustra en la figura 12.11, donde la solución óptima de un problema con una sola variable es $x = 0$ aun cuando la derivada ahí es negativa y no cero. Como este ejemplo tiene una función cóncava para maximizar sujeta a una restricción de no negatividad, que su derivada sea menor o igual a 0 en $x = 0$, es una condición necesaria y suficiente para que $x = 0$ sea óptima.

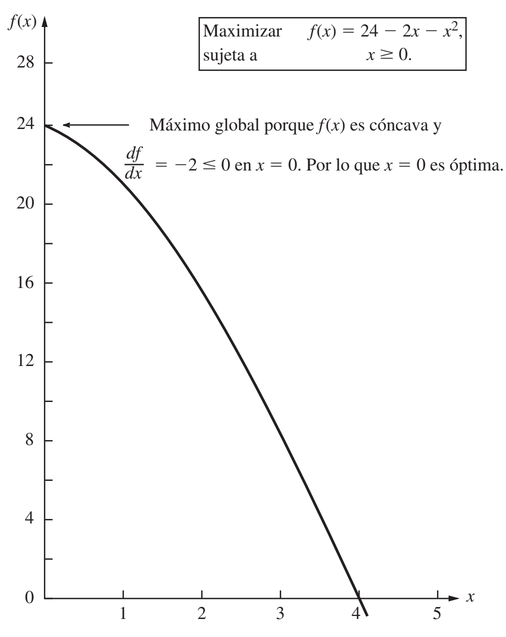


FIGURA 12.11
Ejemplo que ilustra cómo una solución óptima puede estar en un punto en el que la derivada es negativa en lugar de cero, porque ese punto está en la frontera de una restricción de no negatividad.

Un problema que tiene algunas restricciones de no negatividad y que no tiene restricciones funcionales es un caso especial ($m = 0$) de la siguiente clase de problemas.

Optimización restringida linealmente

Los problemas de optimización restringida linealmente se caracterizan por restricciones que se ajustan por completo a la programación lineal, de manera que *todas* las funciones de restricción $g_i(\mathbf{x})$ son lineales, pero la función objetivo $f(\mathbf{x})$ es no lineal. El problema se simplifica de manera notable si sólo se tiene que tomar en cuenta una función no lineal junto con una región factible de programación lineal. Se han desarrollado varios algoritmos especiales basados en una *extensión* del método símplex para analizar la función objetivo no lineal.

Un caso especial importante que se describe a continuación es la programación cuadrática.

Programación cuadrática

De nuevo los problemas de *programación cuadrática* tienen restricciones lineales, pero ahora la función objetivo $f(\mathbf{x})$ debe ser *cuadrática*. Entonces, la única diferencia entre éstos y un problema de programación lineal es que algunos términos de la función objetivo incluyen el *cuadrado* de una variable o el *producto* de dos variables.

Se han desarrollado muchos algoritmos para manejar este caso, con el supuesto adicional de que $f(\mathbf{x})$ es *cóncava*. En la sección 12.7 se presenta un algoritmo que aplica una extensión del método símplex.

La programación cuadrática es muy importante, en parte porque las formulaciones de este tipo surgen de manera natural en muchas aplicaciones. Por ejemplo, el problema de la selección de una cartera con inversiones riesgosas, que se describió en la sección 12.1, se ajusta a este formato. Sin embargo, otra razón por la que es importante es que al resolver problemas generales de optimización restringida linealmente se puede obtener la solución de una sucesión de aproximaciones de programación cuadrática.

Programación convexa

La *programación convexa* abarca una amplia clase de problemas, entre los cuales, como casos especiales, se puede mencionar todos los tipos anteriores cuando $f(\mathbf{x})$ es una función cóncava que debe maximizarse. Los supuestos son:

1. $f(\mathbf{x})$ es cóncava.
2. Cada una de las $g_i(\mathbf{x})$ es convexa.

Como se expuso al final de la sección 12.2, estos supuestos son suficientes para asegurar que un máximo local es un máximo global. (Si, por el contrario, el objetivo fuera *minimizar* $f(\mathbf{x})$, sujeta ya sea a $g_i(\mathbf{x}) \leq b_i$ o a $-g_i(\mathbf{x}) \geq b_i$ para $i = 1, 2, \ldots, m$, el primer supuesto cambiaría a que $f(\mathbf{x})$ debe ser una función *convexa*, puesto que es lo que se requiere para asegurar que un mínimo local sea un mínimo global.) En la sección 12.6 se verá que las condiciones necesarias y suficientes para obtener tal solución óptima son una generalización natural de las condiciones que se acaba de exponer para la *optimización no restringida* y su extensión a la inclusión de *restricciones de no negatividad*. En la sección 12.9 se describen los enfoques algorítmicos para resolver problemas de programación convexa.

Programación separable

La *programación separable* es un caso especial de programación convexa, en donde el supuesto adicional es:

3. Todas las funciones $f(\mathbf{x})$ y $g_i(\mathbf{x})$ son separables.

Una **función separable** es una función en la que *cada término* incluye *una sola variable*, por lo que la función se puede *separar* en una suma de funciones de variables individuales. Por ejemplo, si $f(\mathbf{x})$ es una función separable, se puede expresar como

$$f(\mathbf{x}) = \sum_{j=1}^{n} f_j(x_j),$$

donde cada $f_j(x_j)$ incluye sólo los términos con x_j. En la terminología de programación lineal (vea la sección 3.3), los problemas de programación separable satisfacen los supuestos de aditividad, pero violan el supuesto de proporcionalidad cuando cualquiera de las funciones $f_j(x_j)$ son funciones no lineales.

Para ilustrar, la función objetivo que se considera en la figura 12.6

$$f(x_1, x_2) = 126x_1 - 9x_1^2 + 182x_2 - 13x_2^2$$

es una función separable porque puede ser expresada como

$$f(x_1, x_2) = f_1(x_1) + f_2(x_2)$$

donde $f_1(x_1) = 126x_1 - 9x_1^2$ y $f_2(x_2) = 182x_2 - 13x_2^2$ son cada una funciones de una sola variable x_1 y x_2, respectivamente. Si se usa el mismo razonamiento, se puede verificar que la función considerada en la figura 12.7 también es una función separable.

Es importante distinguir estos problemas de otros de programación convexa, pues cualquier problema de programación separable se puede aproximar muy de cerca mediante uno de programación lineal y, entonces, se puede aplicar el eficiente método símplex. Este enfoque se describe en la sección 12.8. (Para simplificar, se centra la atención en el caso *linealmente restringido*, en el que el tratamiento especial se necesita nada más para la función objetivo.)

Programación no convexa

La *programación no convexa* incluye todos los problemas de programación no lineal que no satisfacen los supuestos de programación convexa. En este caso, aun cuando se tenga éxito en encontrar un *máximo local*, no hay garantía de que sea también un *máximo global*. Por lo tanto, no se cuenta con un algoritmo que garantice encontrar una solución óptima para todos estos problemas; sin embargo, existen algunos algoritmos bastante adecuados para encontrar *máximos locales*, en especial cuando las formas de las funciones no lineales no se desvían demasiado de aquellas que se supuso para programación convexa. En la sección 12.10 se presenta uno de estos algoritmos. En esa misma sección se introducen dos nuevos optimizadores globales (ahora disponibles con LINGO y MPL) para encontrar una solución óptima para problemas de programación no convexa de tamaño moderado, así como un procedimiento de búsqueda que por lo general encontrará una solución cercana a la óptima para problemas aún más grandes.

Ciertos tipos específicos de problemas de programación no convexa se pueden resolver sin mucha dificultad mediante métodos especiales. Dos de ellos, de gran importancia, se presentarán más adelante.

Programación geométrica

Cuando se aplica programación no lineal a problemas de diseño de ingeniería, muchas veces la función objetivo y las funciones de restricción toman la forma

$$g(\mathbf{x}) = \sum_{i=1}^{N} c_i P_i(\mathbf{x}),$$

donde

$$P_i(\mathbf{x}) = x_1^{a_{i1}} x_2^{a_{i2}} \cdots x_n^{a_{in}}, \qquad \text{para } i = 1, 2, \ldots, N.$$

En tales casos, c_i y a_{ij} con frecuencia representan las constantes físicas, mientras que las x_j son las variables de diseño. Estas funciones por lo general no son ni cóncavas ni convexas, por lo que las técnicas de programación convexa no se pueden aplicar en forma directa a estos problemas de *programación geométrica*. Sin embargo, existe un caso importante en el que el problema se puede transformar en un problema de programación convexa equivalente. Este caso es aquel en el que *todos* los coeficientes c_i de cada función son estrictamente positivos, es decir, las funciones son *polinomios positivos generalizados* (ahora llamados **posinomios**), y la función objetivo se tiene

que minimizar. El problema equivalente de programación convexa con variables de decisión y_1, $y_2, \ldots, y_n$ se obtiene al establecer

$$x_j = e^{y_j}, \qquad \text{para } j = 1, 2, \ldots, n$$

en todo el modelo original, de modo que ya se puede aplicar un algoritmo de programación convexa. Se ha desarrollado otro procedimiento de solución para resolver estos problemas de *programación posinomial*, al igual que para problemas de programación geométrica de otros tipos.

Programación fraccional

Suponga que la función objetivo se encuentra en la forma de una *fracción*, esto es, la razón o cociente de dos funciones,

$$\text{Maximizar} \qquad f(\mathbf{x}) = \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})}.$$

Estos problemas de *programación fraccional* surgen, por ejemplo, cuando se maximiza la razón de la producción entre las horas-hombre empleadas (productividad), o la ganancia entre el capital invertido (tasa de rendimiento), o el valor esperado dividido entre la desviación estándar de alguna medida de desempeño de una cartera de inversiones (rendimiento/riesgo). Se han formulado algunos procedimientos de solución especiales para ciertas formas de $f_1(\mathbf{x})$ y $f_2(\mathbf{x})$.

Cuando es posible, el enfoque más directo para resolver un problema de programación fraccional es transformarlo en un problema equivalente de algún tipo estándar que disponga de un procedimiento eficiente. Para ilustrar este enfoque, suponga que $f(\mathbf{x})$ es de la forma de *programación fraccional lineal*

$$f(\mathbf{x}) = \frac{\mathbf{cx} + c_0}{\mathbf{dx} + d_0},$$

donde $\mathbf{c}$ y $\mathbf{d}$ son vectores renglón, $\mathbf{x}$ es un vector columna y c_0 y d_0 son escalares. También suponga que las funciones de restricción $g_i(\mathbf{x})$ son lineales, es decir, las restricciones en forma matricial son $\mathbf{Ax} \leq \mathbf{b}$ y $\mathbf{x} \geq \mathbf{0}$.

Bajo algunos supuestos débiles adicionales, el problema se puede transformar en un problema equivalente de *programación lineal* si se establece

$$\mathbf{y} = \frac{\mathbf{x}}{\mathbf{dx} + d_0} \qquad \text{y} \qquad t = \frac{1}{\mathbf{dx} + d_0},$$

de manera que $\mathbf{x} = \mathbf{y}/t$. Este resultado conduce a

$$\text{Maximizar} \qquad Z = \mathbf{cy} + c_0 t,$$

sujeta a

$$\mathbf{Ay} - \mathbf{b}t \leq \mathbf{0},$$
$$\mathbf{dy} + d_0 t = 1,$$

y

$$\mathbf{y} \geq \mathbf{0}, \qquad t \geq 0,$$

que se puede resolver con el método símplex. En términos generales, se puede usar el mismo tipo de transformación para convertir un problema de programación fraccional con $f_1(\mathbf{x})$ cóncava, $f_2(\mathbf{x})$ convexa y $g_i(\mathbf{x})$ convexa, en un problema equivalente de programación convexa.

Problema de complementariedad

Cuando se estudie la programación cuadrática en la sección 12.7, se verá un ejemplo de cómo la solución de ciertos problemas de programación no lineal se puede reducir a resolver el problema de complementariedad. Dadas las variables $w_1, w_2, \ldots, w_p$ y $z_1, z_2, \ldots, z_p$, el **problema de complementariedad** encuentra una solución factible para el conjunto de restricciones

$$\mathbf{w} = F(\mathbf{z}), \qquad \mathbf{w} \geq \mathbf{0}, \qquad \mathbf{z} \geq \mathbf{0}$$

que también satisface la **restricción de complementariedad**,

$$\mathbf{w}^T\mathbf{z} = 0.$$

Aquí, **w** y **z** son vectores columna, F es una función dada con valores vectoriales y el superíndice T denota la transpuesta (vea el apéndice 4). El problema no tiene función objetivo, de manera que, desde un punto de vista técnico, no es un problema de programación no lineal completo. Se llama problema de complementariedad por las relaciones complementarias que establecen que

$$w_i = 0 \quad \text{o} \quad z_i = 0 \quad \text{(o ambas)} \quad \text{para cada } i = 1, 2, \ldots, p.$$

Un caso en especial importante es el **problema de complementariedad lineal**, donde

$$F(\mathbf{z}) = \mathbf{q} + \mathbf{M}\mathbf{z},$$

donde **q** es un vector columna dado y **M** es una matriz dada de orden $p \times p$. Se dispone de algoritmos eficientes para resolver este problema bajo algunos supuestos adecuados[6] sobre las propiedades de la matriz **M**. Uno de éstos requiere pivotear de una solución básica factible (BF) a la siguiente, en forma muy parecida a la del método símplex para programación lineal.

Además de tener aplicaciones en programación no lineal, los problemas de complementariedad se utilizan en teoría de juegos, problemas de equilibrio económico y problemas de equilibrio en ingeniería.

12.4 OPTIMIZACIÓN NO RESTRINGIDA DE UNA VARIABLE

A continuación se explicará cómo se pueden resolver algunos tipos de problemas como los que se acaban de describir, comenzando por el caso más sencillo, la *optimización no restringida* con una sola variable x ($n = 1$), donde la función diferenciable $f(x)$ que debe maximizarse es *cóncava*.[7] En consecuencia, la *condición necesaria* y *suficiente* para que una solución particular $x = x^*$ sea óptima (un máximo global) es

$$\frac{df}{dx} = 0 \qquad \text{en } x = x^*,$$

como se ilustra en la figura 12.12. Si en esta ecuación se puede despejar x^* de modo directo, el problema llega a su fin; pero si $f(x)$ no es una función sencilla y su derivada no es una función lineal o cuadrática, tal vez sea imposible resolver la ecuación *analítica*. De ser así, existe una cantidad de *procedimientos de búsqueda* para resolver el problema en forma *numérica*.

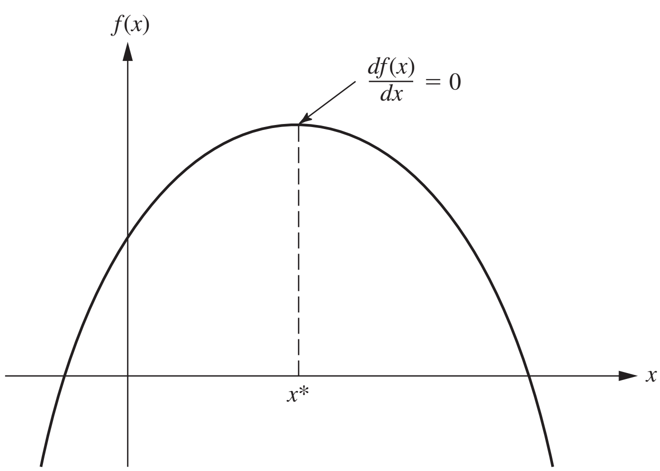


FIGURA 12.12
Problema de programación no restringida de una variable cuando la función es cóncava.

[6] Vea R. W. Cottle, J.-S. Pang y R. E. Stone, *The Linear Complementarity Problem*, Academic Press, Boston, 1992.

[7] Vea el comienzo del apéndice 3 donde podrá encontrar una revisión del caso correspondiente cuando $f(x)$ no es cóncava.

El enfoque con cualquiera de estos procedimientos de búsqueda implica encontrar una serie de *ensayos de solución* que conduzcan hacia una solución óptima. En cada iteración se comienza con la solución de prueba actual para llevar a cabo una búsqueda sistemática, que culmina con la identificación de una nueva solución de prueba *mejorada*. El procedimiento continúa hasta que la solución de prueba haya convergido hacia una solución óptima, en caso de que exista una.

A continuación se describirán dos procedimientos de búsqueda comunes. El primero —*método de bisección*— se escogió porque es un procedimiento intuitivo y directo. El segundo —*método de Newton*— se incluyó porque tiene un papel fundamental en la programación no lineal en general.

Método de bisección

Este procedimiento de búsqueda siempre se puede aplicar cuando $f(x)$ es cóncava —de forma que la segunda derivada sea negativa o cero para toda x— como se mostró en la figura 12.12. Este método puede usarse también para algunas otras funciones. En particular, si x^* denota la solución óptima, todo lo que se necesita[8] es que

$$\frac{df(x)}{dx} > 0 \quad \text{si } x < x^*,$$

$$\frac{df(x)}{dx} = 0 \quad \text{si } x = x^*,$$

$$\frac{df(x)}{dx} < 0 \quad \text{si } x > x^*.$$

Estas condiciones se cumplen de manera automática si $f(x)$ es cóncava, pero también se pueden cumplir cuando la segunda derivada es positiva para algunos (pero no todos) los valores de x.

La idea que subyace al método de bisección es muy intuitiva, esto es, el hecho de que la pendiente (derivada) sea positiva o negativa en una solución de prueba indica de forma definitiva si la mejoría respectiva cae de inmediato a la derecha o a la izquierda. De esta forma, si la derivada evaluada en un valor particular de x es positiva, entonces x^* debe ser *mayor* que dicha x (vea la figura 12.12), de manera que ésta se convierte en una *cota inferior* para las soluciones de prueba que deben considerarse de este punto en adelante. De manera inversa, si la derivada es *negativa*, entonces x^* debe ser *menor* que esta x, por lo cual ésta se convertiría en una *cota superior*. Por tanto, después de que se han identificado ambos tipos de cotas, cada nueva solución de prueba seleccionada entre las cotas actuales proporciona un intervalo más estrecho de alguno de los dos tipos, con lo que reduce la búsqueda de ahí en adelante. Cuando se usa una regla razonable para seleccionar cada solución de prueba de esta forma, la *secuencia* resultante de soluciones de prueba debe *converger* a x^*. En la práctica, esto significa continuar la secuencia hasta que la distancia entre las cotas sea tan pequeña como para que la siguiente solución de prueba esté dentro de una *tolerancia de error* para x^* especificada con anterioridad.

Este proceso completo se resume a continuación, dada la notación

x' = solución de prueba actual,

$\underline{x}$ = cota inferior actual para x^*,

$\overline{x}$ = cota superior actual para x^*,

ϵ = tolerancia del error de x^*.

Aunque existen algunas reglas razonables para seleccionar cada nueva solución de prueba, la que se utiliza en el método de bisección es la regla del **punto medio** —llamada en forma tradicional *plan de búsqueda de Bolzano*—, la cual dice simplemente que se debe seleccionar el punto medio entre las dos cotas actuales.

[8] Otra posibilidad es que la parte superior de $f(x)$ sea plana de manera que x sea óptima en algún intervalo $[a, b]$. En este caso, el procedimiento aún convergirá a una de estas soluciones óptimas tan pronto como la derivada sea positiva para $x < a$ y negativa para $x > b$.

Resumen del método de bisección

Paso inicial: Seleccione ϵ. Encuentre la $\underline{x}$ y $\overline{x}$ iniciales por inspección o mediante la ubicación de cualesquiera valores respectivos de x en los cuales la derivada sea positiva y negativa. Seleccione una solución de prueba inicial

$$x' = \frac{\underline{x} + \overline{x}}{2}.$$

Iteración:

1. Evalúe $\dfrac{df(x)}{dx}$ para $x = x'$.

2. Si $\dfrac{df(x)}{dx} \geq 0$, redefina $\underline{x} = x'$.

3. Si $\dfrac{df(x)}{dx} \leq 0$, redefina $\overline{x} = x'$.

4. Seleccione una nueva $x' = \dfrac{\underline{x} + \overline{x}}{2}$.

Regla de detención: Detenerse si $\overline{x} - \underline{x} \leq 2\epsilon$, de modo que la nueva x' deba estar dentro del ϵ para x^*. En caso contrario debe realizarse otra iteración.

Ahora se ilustrará el método de bisección mediante su aplicación al siguiente ejemplo.

Ejemplo. Suponga que la función que debe maximizarse es

$$f(x) = 12x - 3x^4 - 2x^6,$$

como se grafica en la figura 12.13. Sus primeras dos derivadas son

$$\frac{df(x)}{dx} = 12(1 - x^3 - x^5),$$

$$\frac{d^2f(x)}{dx^2} = -12(3x^2 + 5x^4).$$

Debido a que la segunda derivada es no positiva en todas partes, $f(x)$ es una función cóncava, por lo que el método de bisección puede aplicarse con seguridad para encontrar su máximo global (en el supuesto de que éste exista).

FIGURA 12.13
Ejemplo del método de bisección.

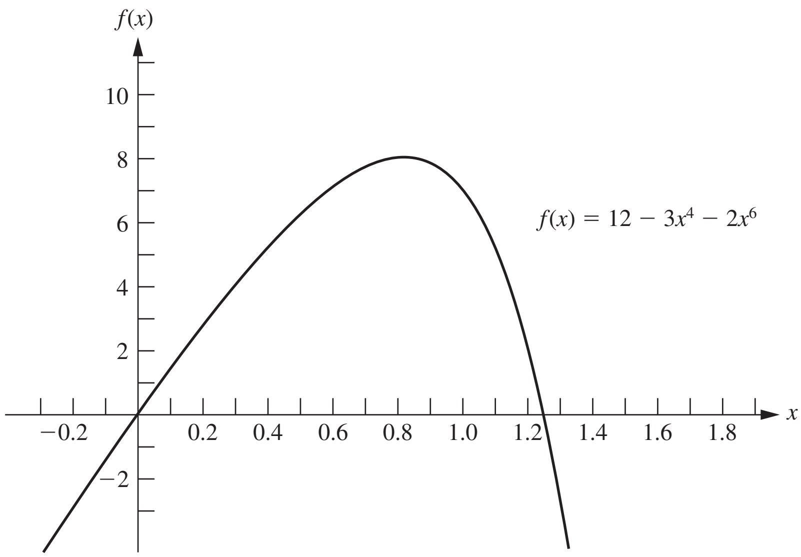


TABLA 12.1 Aplicación del método de bisección al ejemplo

Iteración	$\frac{df(x)}{dx}$	$\underline{x}$	$\bar{x}$	Nueva x'	$f(x')$
0		0	2	1	7.0000
1	−12	0	1	0.5	5.7812
2	+10.12	0.5	1	0.75	7.6948
3	+4.09	0.75	1	0.875	7.8439
4	−2.19	0.75	0.875	0.8125	7.8672
5	+1.31	0.8125	0.875	0.84375	7.8829
6	−0.34	0.8125	0.84375	0.828125	7.8815
7	+0.51	0.828125	0.84375	0.8359375	7.8839
Detener					

Una inspección rápida de esta función —sin siquiera construir su gráfica como se muestra en la figura 12.13— indica que $f(x)$ es positiva para valores positivos pequeños de x, pero es negativa para $x < 0$ o $x > 2$. Por lo tanto, $\underline{x} = 0$ y $\bar{x} = 2$ se pueden usar como las cotas iniciales, con su punto medio, $x' = 1$ como la solución de prueba inicial. Sea $\epsilon = 0.01$ la tolerancia al error para x^* en la regla de detención, entonces la $(\bar{x} - \underline{x})$ final es ≤ 0.02 con la x' final en el punto medio.

En consecuencia, la aplicación del método de bisección produce la secuencia de resultados que se muestra en la tabla 12.1. [Esta tabla incluye tanto los valores de la función como los de la derivada para su información, donde la derivada se evalúa en la solución de prueba generada en la iteración *precedente*. Sin embargo, observe que en realidad el algoritmo no necesita calcular $f(x')$ y que sólo se requiere calcular la derivada lo suficiente para determinar su signo.] La conclusión es que

$$x^* \approx 0.836,$$
$$0.828125 < x^* < 0.84375.$$

En el IOR Tutorial se incluye una rutina interactiva para ejecutar el método de bisección.

Método de Newton

Aunque el método de bisección es un procedimiento intuitivo y directo, tiene la desventaja de que converge de una manera relativamente lenta hacia una solución óptima. Cada iteración sólo disminuye la diferencia entre las cotas a la mitad. Por lo tanto, aun con la función tan simple que se consideró en la tabla 12.1, se requirieron siete iteraciones para reducir la tolerancia al error de x^* a menos de 0.01. Para reducir esta tolerancia al error a menos de 0.0001 se requerirían otras siete iteraciones.

La razón básica de esta convergencia lenta es que la única información que se ha usado sobre $f(x)$ es el valor de la primera derivada $f'(x)$ en los respectivos valores de prueba de x. Si se considerara también la segunda derivada $f''(x)$, se podría obtener información adicional útil. Este enfoque es el que aplica el *método de Newton*.[9]

La idea básica detrás del método de Newton es aproximar $f(x)$ a la vecindad de la solución de prueba inicial mediante una función cuadrática y después maximizar (o minimizar) la función aproximada exactamente para obtener la nueva solución de prueba y así iniciar la siguiente iteración. (Esta idea de trabajar con una **aproximación cuadrática** de la función objetivo se ha convertido en una característica clave de muchos algoritmos para tipos más generales de problemas de programación lineal.) Esta función cuadrática de aproximación se obtiene al truncar la serie de Taylor después del término de la segunda derivada. En particular, se considera que x_{i+1} es la so-

[9] Este método se debe al gran matemático y físico del siglo XVII, Sir Isaac Newton. Cuando aún era un joven estudiante de la Universidad de Cambridge (Inglaterra), Newton obtuvo ventaja de que la universidad estuvo cerrada por dos años (debido a la peste bubónica que devastó Europa en 1664-1665) para descubrir la ley de la gravitación universal e inventar el cálculo (entre otros logros). Su desarrollo del cálculo condujo a este método.

lución de prueba generada en la iteración i para iniciar la iteración $i + 1$ —de manera que x_1 es la solución de prueba que proporciona el usuario para comenzar la iteración 1—, por lo que la serie de Taylor truncada para x_{i+1} es

$$f(x_{i+1}) \approx f(x_i) + f'(x_i)(x_{i+1} - x_i) + \frac{f''(x_i)}{2}(x_{i+1} - x_i)^2.$$

Al haber fijado x_i al inicio de la iteración i, observe que $f(x_i)$, $f'(x_i)$ y $f''(x_i)$ también son constantes fijas en esta función de aproximación a la derecha. Entonces, esta función de aproximación es sólo una función cuadrática de x_{i1}. Aún más, esta función cuadrática es una aproximación tan buena de $f(x_{i+1})$ en la vecindad de x_i, que sus valores y los de su primera y segunda derivadas son exactamente los mismos cuando $x_{i+1} = x_i$.

Ahora esta función cuadrática se puede maximizar de la manera usual al igualar su primera derivada a cero y al resolver para x_{i+1}. (Recuerde que se supone que $f(x)$ es cóncava, lo cual implica que esta función cuadrática es cóncava, por lo cual cuando se iguala la primera derivada a cero la solución será un máximo global.) Esta primera derivada es

$$f'(x_{i+1}) \approx f'(x_i) + f''(x_i)(x_{i+1} - x_i)$$

como x_i, $f(x_i)$, $f'(x_i)$ y $f''(x_i)$ son constantes. Al igualar a cero la primera derivada de la derecha se obtiene

$$f'(x_{i+1}) + f''(x_i)(x_{i+1} - x_i) = 0,$$

lo cual directamente conduce de manera algebraica a la solución,

$$x_{i+1} = x_i - \frac{f'(x_i)}{f''(x_i)}.$$

Ésta es la fórmula clave que se utiliza en cada iteración i para calcular la siguiente solución de prueba x_{i+1} después de obtener la solución de prueba x_i para iniciar la iteración i y después calcular la primera y segunda derivadas en x_i. (Cuando se minimiza una función convexa se emplea la misma fórmula.)

Las iteraciones que generan nuevas soluciones de prueba de esta forma deben continuar hasta que éstas hayan convergido en esencia. Un criterio para sostener la convergencia es que $|x_{i+1} - x_i|$ se haya empequeñecido lo suficiente. Otro es que $f'(x)$ esté suficientemente cerca de cero. Y uno más es que $|f(x_{i+1}) - f(x_i)|$ sea suficientemente pequeño. Si se elige el primer criterio, se define ϵ como un valor tal que el algoritmo se detenga cuando $|x_{i+1} - x_i| \leq \epsilon$.

Aquí se presenta la descripción completa del algoritmo.

Resumen del método de Newton

Paso inicial: Seleccione ϵ. Encuentre una solución de prueba inicial x_i por inspección. Establezca $i = 1$.

Iteración i:

1. Calcule $f'(x_i)$ y $f''(x_i)$. [El cálculo de $f(x_i)$ es opcional.]
2. Establezca $x_{i+1} = x_i - \dfrac{f'(x_i)}{f''(x_i)}$.

Regla de detención: Si $|x_{i+1} - x_i| \leq \epsilon$, , debe detenerse; en esencia x_{i+1} es la solución óptima. De otra manera, se redefine $i = i + 1$ y se realiza otra iteración.

TABLA 12.2 Aplicación del método de Newton al ejemplo

Iteración i	x_i	$f(x_i)$	$f'(x_i)$	$f''(x_i)$	x_{i+1}
1	1	7	−12	−96	0.875
2	0.875	7.8439	−2.1940	−62.733	0.84003
3	0.84003	7.8838	−0.1325	−55.279	0.83763
4	0.83763	7.8839	−0.0006	−54.790	0.83762

Ejemplo. Ahora se aplicará el método de Newton al mismo ejemplo al que se aplicó el método de bisección. Como se mostró en la figura 12.13, la función que debe maximizarse es

$$f(x) = 12x - 3x^4 - 2x^6.$$

En consecuencia, la fórmula para calcular la nueva solución de prueba (x_{i+1}) a partir de la actual (x_i) es

$$x_{i+1} = x_i - \frac{f'(x_i)}{f''(x_i)} = x_i - \frac{12(1 - x^3 - x^5)}{-12(3x^2 + 5x^4)} = x_i + \frac{1 - x^3 - x^5}{3x^2 + 5x^4}.$$

Después de seleccionar $\epsilon = 0.00001$ y de elegir $x_1 = 1$ como la solución de prueba inicial, en la tabla 12.2 se muestran los resultados de aplicar el método de Newton a este ejemplo. Después de sólo cuatro iteraciones, este método ha convergido a $x = 0.83762$ como la solución óptima, con un grado muy alto de precisión.

Una comparación de esta tabla con la tabla 12.1 ilustra cuánto más rápido converge el método de Newton que el método de bisección. Si se aplica este último, se requerirían cerca de 20 iteraciones para converger con el mismo grado de precisión que logró el método de Newton después de sólo cuatro iteraciones.

Aunque esta convergencia rápida es muy usual en el método de Newton, su desempeño varía de problema a problema. Debido a que este enfoque se basa en el uso de una aproximación cuadrática de $f(x)$, su desempeño es afectado por el grado de precisión de la aproximación.

12.5 OPTIMIZACIÓN NO RESTRINGIDA DE VARIAS VARIABLES

Ahora considere el problema de maximizar una función *cóncava* $f(\mathbf{x})$ de variables *múltiples*, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, en la que no existen restricciones sobre los valores factibles. Suponga de nuevo que la condición necesaria y suficiente para la optimalidad, dada por el sistema de ecuaciones que se obtiene al establecer las respectivas derivadas parciales iguales a cero (vea la sección 12.3), no se puede resolver en forma analítica, por lo que debe emplearse un procedimiento de búsqueda numérico.

Igual que en el caso de una variable, existen varios procedimientos de búsqueda para resolver este tipo de problema en forma numérica. Uno de éstos —*el procedimiento de búsqueda del gradiente*— es muy importante porque identifica y utiliza la dirección de movimiento, desde la solución de prueba actual, que maximiza la tasa a la cual se incrementa $f(\mathbf{x})$. Ésta es una de las ideas clave de la programación no lineal. Las adaptaciones de esta idea para tomar en cuenta las restricciones también son una característica central de muchos algoritmos que se utilizan para llevar a cabo una optimización *restringida*.

Después de analizar este procedimiento con algún nivel de detalle, se describirá en forma breve cómo se extiende el método de Newton al caso de las variables múltiples.

Procedimiento de búsqueda de gradiente

En la sección 12.4 se usó el valor de la derivada ordinaria mediante el método de bisección para elegir una de sólo dos direcciones posibles (aumentar x o disminuir x) para pasar de la solución de prueba actual a la siguiente. La meta era alcanzar, en algún momento, un punto en el que la derivada fuera (en esencia) cero. Ahora se tienen *innumerables* direcciones posibles hacia dónde moverse; corresponden a las *tasas proporcionales* posibles a las cuales las respectivas variables pueden cambiar. La meta es alcanzar, finalmente, un punto en el que todas las derivadas parciales sean —en esencia— 0. Por tanto, la extensión del procedimiento de búsqueda en una dimensión requiere emplear los valores de las derivadas parciales para seleccionar la dirección específica en la que conviene moverse. Esta selección implica el uso del gradiente de la función objetivo, como se describirá en seguida.

Como se supone que la función objetivo $f(\mathbf{x})$ es diferenciable, posee un gradiente denotado por $\nabla f(\mathbf{x})$ en cada punto $\mathbf{x}$. En particular, el **gradiente** en un punto específico $\mathbf{x} = \mathbf{x}'$ es el *vector* cuyos elementos son las *derivadas parciales* respectivas evaluadas en $\mathbf{x} = \mathbf{x}'$, por lo cual

$$\nabla f(\mathbf{x}') = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\right) \qquad \text{en } \mathbf{x} = \mathbf{x}'.$$

El significado del gradiente es que el cambio (infinitesimal) de $\mathbf{x}$, que *maximiza* la tasa a la que $f(\mathbf{x})$ aumenta, es el cambio que es *proporcional* a $\nabla f(\mathbf{x})$. Para expresar en forma geométrica esta idea, la "dirección" del gradiente, $\nabla f(\mathbf{x}')$, se interpreta como la *dirección* del segmento de recta dirigido (flecha) que va del origen (0, 0, . . ., 0) al punto $(\partial f/\partial x_1, \partial f/\partial x_2, \ldots, \partial f/\partial x_n)$, donde $\partial f/\partial x_j$ se evalúa en $x_j = x_j'$. En consecuencia, se puede decir que la tasa a la que aumenta $f(\mathbf{x})$ se maximiza si los cambios (infinitesimales) de $\mathbf{x}$ se hacen en la *dirección* del gradiente $\nabla f(\mathbf{x})$. Como el objetivo es encontrar la solución factible que maximice $f(\mathbf{x})$, parece adecuado intentar moverse lo más posible en la dirección del gradiente.

Debido a que el problema en cuestión no tiene restricciones, esta interpretación del gradiente sugiere que un procedimiento de búsqueda eficiente debe moverse en la dirección del gradiente hasta que (en esencia) alcance una solución óptima $\mathbf{x}^*$, en la que $\nabla f(\mathbf{x}^*) = \mathbf{0}$. Sin embargo, no resultaría práctico cambiar $\mathbf{x}$ de manera *continua* en la dirección de $\nabla f(\mathbf{x})$, puesto que esta serie de cambios requeriría una *reevaluación* continua de $\partial f/\partial x_j$ y el cambio de dirección de la trayectoria. Entonces, una mejor forma de hacerlo es continuar el movimiento en una dirección *fija* a partir de la solución de prueba actual, sin detenerse, hasta que $f(\mathbf{x})$ deje de crecer. Este punto de detención sería la siguiente solución de prueba, por lo que se debe volver a calcular el gradiente para determinar la nueva dirección de movimiento. Con este enfoque, cada iteración incluye cambiar la solución de prueba actual $\mathbf{x}'$ de la siguiente manera:

$$\text{Se modifica} \quad \boldsymbol{x}' = \boldsymbol{x}' + t^* \nabla f(\boldsymbol{x}'),$$

donde t^* es el valor positivo de t que *maximiza* $f(\mathbf{x}' + t\,\nabla f(\mathbf{x}'))$; es decir,

$$f(\mathbf{x}' + t^*\,\nabla f(\mathbf{x}')) = \max_{t \geq 0} f(\mathbf{x}' + t\,\nabla f(\mathbf{x}')).$$

[Note que $f(\mathbf{x}' + t\,\nabla f(\mathbf{x}'))$ es sencillamente $f(\mathbf{x})$ donde

$$x_j = x_j' + t\left(\frac{\partial f}{\partial x_j}\right)_{\mathbf{x}=\mathbf{x}'}, \qquad \text{para } j = 1, 2, \ldots, n,$$

y estas expresiones de x_j incluyen sólo constantes y t, de manera que $f(\mathbf{x})$ se convierte en una función de una sola variable t.] Las iteraciones de este procedimiento de búsqueda del gradiente continúan hasta que $\nabla f(\mathbf{x}) = 0$ dentro de una pequeña tolerancia ϵ, o sea, hasta que

$$\left|\frac{\partial f}{\partial x_j}\right| \leq \epsilon \qquad \text{para } j = 1, 2, \ldots, n.$$[10]

Una analogía puede ayudar a aclarar este procedimiento. Suponga que una persona quiere subir a la cumbre de una montaña. Esta persona es miope y no puede ver la cumbre para caminar en esa dirección, pero cuando se detiene, puede ver el piso a su alrededor y determinar la dirección en la que la pendiente de la montaña es más pronunciada. La persona puede caminar en línea recta. Mientras camina, también es capaz de percibir cuándo ya no va hacia arriba (pendiente cero en esa dirección). Bajo el supuesto de que la montaña es *cóncava*, el individuo puede usar el *procedimiento de búsqueda del gradiente* para escalar de manera eficiente hasta la cima. Este problema tiene *dos variables*, donde (x_1, x_2) representa las coordenadas (se hace caso omiso de la altura) de la localización actual. La función $f(x_1, x_2)$ da la altura de la montaña en (x_1, x_2). La persona comienza cada iteración en el lugar en que se encuentra (solución de prueba actual) y determina la

[10] En general, esta regla de detención proporcionará una solución $\mathbf{x}$ cercana a una solución óptima $\mathbf{x}^*$, con un valor de $f(\mathbf{x})$ muy cercano a $f(\mathbf{x}^*)$. Sin embargo, esto no se puede garantizar puesto que es posible que la función mantenga una pendiente positiva muy pequeña ($\leq \epsilon$) para una distancia grande entre $\mathbf{x}$ y $\mathbf{x}^*$.

dirección [en el sistema de coordenadas (x_1, x_2)] en la que la montaña tiene la mayor pendiente (la dirección del gradiente) en ese punto. Después, comienza a caminar en esa dirección fija y continúa haciéndolo mientras sigue ascendiendo. Luego se detiene en un nuevo punto de prueba (solución) cuando la montaña se nivela en la dirección en que camina; en este punto se prepara para realizar otra iteración en otra dirección. Estas iteraciones continúan, en una trayectoria en zigzag hacia arriba, hasta que alcanza un punto de prueba en el que la pendiente es en esencia cero en todas direcciones. Bajo el supuesto de que la montaña $[f(x_1, x_2)]$ es cóncava, en principio, la persona debe estar en la cima de la montaña.

Por lo general, la parte más difícil del procedimiento de búsqueda del gradiente es encontrar t^*, el valor de t que maximiza f en la dirección del gradiente, en cada iteración. Como $\mathbf{x}$ y $\nabla f(\mathbf{x})$ tienen valores fijos para la maximización y como $f(\mathbf{x})$ es cóncava, este problema se debe ver como el de maximizar una función *cóncava de una sola variable* t. En efecto, se puede resolver con el tipo de procedimientos de búsqueda de la optimización no restringida de una sola variable descritos en la sección 12.4 (donde la cota inferior inicial sobre t debe ser no negativa por la restricción de $t \geq 0$). De otra manera, si f es una función simple, es posible que se pueda obtener una solución analítica al establecer la derivada con respecto a t igual a cero para después resolver.

Resumen del procedimiento de búsqueda del gradiente

Paso inicial: Elija ϵ y cualquier solución de prueba inicial x'. Vaya a la regla de detención.

Iteración:

1. Exprese $f(\mathbf{x}' + t\,\nabla f(\mathbf{x}'))$ como una función de t al establecer

$$x_j = x_j' + t\left(\frac{\partial f}{\partial x_j}\right)_{\mathbf{x}=\mathbf{x}'}, \qquad \text{para } j = 1, 2, \ldots, n,$$

y después sustituya estas expresiones en $f(\mathbf{x})$.

2. Utilice el procedimiento de búsqueda en una dimensión (o cálculo) para encontrar $t = t^*$ que maximiza $f(\mathbf{x}' + t\,\nabla f(\mathbf{x}'))$ para $t \geq 0$.

3. Establezca $\mathbf{x}' = \mathbf{x}' + t^*\,\nabla f(\mathbf{x}')$. Después, pase a la regla de detención.

Regla de detención: Evalúe $\nabla f(\mathbf{x}')$ en $\mathbf{x} = \mathbf{x}'$. Verifique si

$$\left|\frac{\partial f}{\partial x_j}\right| \leq \epsilon \qquad \text{para toda } j = 1, 2, \ldots, n.$$

Si es así, el proceso se detiene con la $\mathbf{x}'$ actual como la aproximación a una solución óptima $\mathbf{x}^*$ deseada. De otra manera, se realiza otra iteración.

Ahora se ilustrará este procedimiento.

Ejemplo. Considere el siguiente problema de dos variables:

$$\text{Maximizar} \qquad f(\mathbf{x}) = 2x_1x_2 + 2x_2 - x_1^2 - 2x_2^2.$$

Entonces,

$$\frac{\partial f}{\partial x_1} = 2x_2 - 2x_1,$$

$$\frac{\partial f}{\partial x_2} = 2x_1 + 2 - 4x_2.$$

También se puede verificar (vea el apéndice 2) que $f(\mathbf{x})$ es cóncava.

Para comenzar el procedimiento de búsqueda del gradiente ϵ, suponga que se elige $\mathbf{x} = (0, 0)$ como solución de prueba inicial. Como en este punto las respectivas derivadas parciales son 0 y 2, el gradiente es

$$\nabla f(0, 0) = (0, 2).$$

Con $\epsilon < 2$, la regla de detención indica la realización de una iteración.

Iteración 1: Con los valores de 0 y 2 para las derivadas parciales respectivas, la primera iteración inicia al establecer

$$x_1 = 0 + t(0) = 0,$$
$$x_2 = 0 + t(2) = 2t,$$

y después se sustituyen estas expresiones en $f(\mathbf{x})$ para obtener

$$\begin{aligned} f(\mathbf{x}' + t\,\nabla f(\mathbf{x}')) &= f(0, 2t) \\ &= 2(0)(2t) + 2(2t) - 0^2 - 2(2t)^2 \\ &= 4t - 8t^2. \end{aligned}$$

Debido a que

$$f(0, 2t^*) = \max_{t \geq 0} f(0, 2t) = \max_{t \geq 0} \{4t - 8t^2\}$$

y

$$\frac{d}{dt}(4t - 8t^2) = 4 - 16t = 0,$$

se deduce que

$$t^* = \frac{1}{4},$$

así,

Se redefine $\quad \mathbf{x}' = (0, 0) + \frac{1}{4}(0, 2) = \left(0, \frac{1}{2}\right).$

Con esto termina la primera iteración. El gradiente de esta nueva solución de prueba es

$$\nabla f\left(0, \frac{1}{2}\right) = (1, 0).$$

Con $\epsilon < 1$, ahora la regla de detención indica que debe realizarse otra iteración.

Iteración 2: Para comenzar con la segunda iteración, se usan los valores de 1 y 0 para las derivadas parciales respectivas y se establece

$$\mathbf{x} = \left(0, \frac{1}{2}\right) + t(1, 0) = \left(t, \frac{1}{2}\right),$$

así,

$$\begin{aligned} f(\mathbf{x}' + t\,\nabla f(\mathbf{x}')) &= f\left(0 + t, \frac{1}{2} + 0t\right) = f\left(t, \frac{1}{2}\right) \\ &= (2t)\left(\frac{1}{2}\right) + 2\left(\frac{1}{2}\right) - t^2 - 2\left(\frac{1}{2}\right)^2 \\ &= t - t^2 + \frac{1}{2}. \end{aligned}$$

Debido a que

$$f\left(t^*, \frac{1}{2}\right) = \max_{t \geq 0} f\left(t, \frac{1}{2}\right) = \max_{t \geq 0} \left\{t - t^2 + \frac{1}{2}\right\}$$

y

$$\frac{d}{dt}\left(t - t^2 + \frac{1}{2}\right) = 1 - 2t = 0,$$

TABLA 12.3 Aplicación del procedimiento de búsqueda del gradiente al ejemplo

Iteración	$\mathbf{x}'$	$\nabla f(\mathbf{x}')$	$\mathbf{x}' + t\,\nabla f(\mathbf{x}')$	$f(\mathbf{x}' + t\,\nabla f(\mathbf{x}'))$	t^*	$\mathbf{x}' + t^*\,\nabla f(\mathbf{x}')$
1	$(0, 0)$	$(0, 2)$	$(0, 2t)$	$4t - 8t^2$	$\frac{1}{4}$	$\left(0, \frac{1}{2}\right)$
2	$\left(0, \frac{1}{2}\right)$	$(1, 0)$	$\left(t, \frac{1}{2}\right)$	$t - t^2 + \frac{1}{2}$	$\frac{1}{2}$	$\left(\frac{1}{2}, \frac{1}{2}\right)$

entonces

$$t^* = \frac{1}{2},$$

por lo que

$$\text{Se redefine} \qquad \mathbf{x}' = \left(0, \frac{1}{2}\right) + \frac{1}{2}(1, 0) = \left(\frac{1}{2}, \frac{1}{2}\right).$$

Con esta igualdad se completa la segunda iteración. Con un valor de ϵ típicamente pequeño, el procedimiento ahora continuaría por algunas iteraciones más de una manera similar. (Se omitirán los detalles.)

Una manera sencilla de organizar este trabajo es escribir una tabla como la 12.3, que resume las dos iteraciones anteriores. En cada iteración, la segunda columna muestra la solución de prueba actual y la última muestra la nueva solución de prueba eventual, que después se escribe abajo, en la segunda columna, para la siguiente iteración. La cuarta columna proporciona las expresiones para x_j en términos de t, que se deben sustituir en $f(\mathbf{x})$ para dar la quinta columna.

Si se continúa de esta manera, las soluciones de prueba subsecuentes serán $(\frac{1}{2}, \frac{3}{4})$, $(\frac{3}{4}, \frac{3}{4})$, $(\frac{3}{4}, \frac{7}{8})$, $(\frac{7}{8}, \frac{7}{8})$, . . . , como se muestra en la figura 12.14. Dado que estos puntos convergen a $\mathbf{x}^* = (1, 1)$, ésta es la solución óptima, como se puede verificar por el hecho de que

$$\nabla f(1, 1) = (0, 0).$$

Sin embargo, como esta sucesión convergente de soluciones de prueba nunca alcanza su límite, en realidad el procedimiento se detendrá en algún punto (dependiendo de ϵ) un poco antes de (1, 1) como aproximación final de $\mathbf{x}^*$.

En la figura 12.14 se sugiere que el procedimiento de búsqueda del gradiente marca una trayectoria en zigzag hacia la solución óptima en lugar de moverse en línea recta. Si se toma en cuenta este comportamiento, se pueden realizar algunas modificaciones que *aceleren* el movimiento hacia el óptimo.

Si $f(\mathbf{x})$ *no* fuera una función cóncava, de todas maneras este procedimiento convergiría a un máximo *local*. El único cambio en la descripción del procedimiento para este caso es que t^* ahora corresponde al *primer máximo local* de $f(\mathbf{x}' + t\,\nabla f(\mathbf{x}'))$ conforme t aumenta su valor desde 0.

Si se quisiera *minimizar* la función objetivo $f(\mathbf{x})$, un cambio en el procedimiento sería moverse, en cada iteración, en la dirección *opuesta* al gradiente; en otras palabras, la regla para obtener el siguiente punto sería

$$\text{Redefinir} \quad \mathbf{x}' = \mathbf{x}' - t^*\,\nabla f(\mathbf{x}').$$

El otro único cambio es que t^* ahora sería el valor negativo de la t que *minimiza* $f(\mathbf{x}' - t\,\nabla f(\mathbf{x}'))$; esto es,

$$f(\mathbf{x}' - t^*\,\nabla f(\mathbf{x}')) = \min_{t \geq 0} f(\mathbf{x}' - t\,\nabla f(\mathbf{x}')).$$

Tanto en la sección Worked Examples como en el OR Tutor del CD-ROM se incluyen **ejemplos adicionales** de la aplicación del procedimiento de búsqueda del gradiente. En el IOR Tutorial se incluye una rutina interactiva y una automática para aplicar este algoritmo.

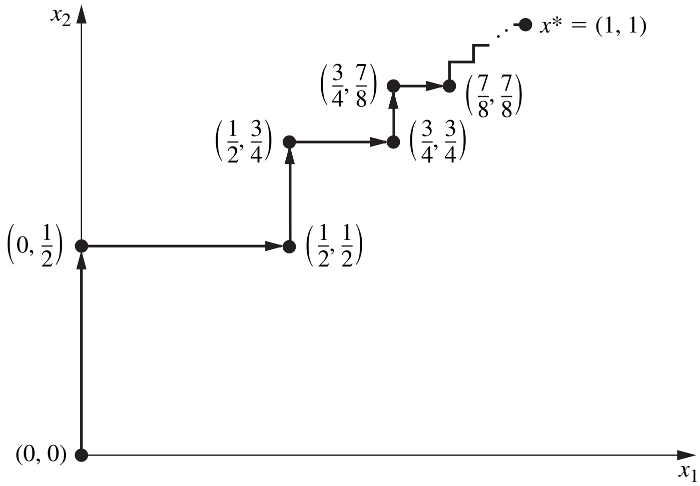


FIGURA 12.14
Ilustración del procedimiento de búsqueda del gradiente cuando $f(x_1, x_2) = 2x_1x_2 + 2x_2 - x_1^2 - 2x_2^2$.

Método de Newton

En la sección 12.4 se describe cómo se puede usar el método de Newton para resolver problemas de optimización restringida con una *sola variable*. En realidad, la versión general del método de Newton está diseñada para resolver problemas de optimización restringida con *variables múltiples*. La idea básica es la misma que se describió en la sección 12.4, es decir, se trabaja con una aproximación cuadrática de la función objetivo $f(\mathbf{x})$, donde, en este caso, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Esta *función cuadrática de aproximación* se obtiene al truncar la serie de Taylor alrededor de la solución de prueba actual después del término de la segunda derivada. Después, esta función aproximada se maximiza (o minimiza) exactamente para obtener la nueva solución de prueba para iniciar la siguiente iteración.

Cuando la función objetivo es cóncava y tanto la solución de prueba actual **x** como su gradiente $\nabla f(\mathbf{x})$ se escriben como *vectores columna*, la solución $\mathbf{x}'$ que maximiza la función cuadrática de aproximación tiene la forma

$$\mathbf{x}' = \mathbf{x} - \left[\nabla^2 f(\mathbf{x})\right]^{-1} \nabla f(\mathbf{x}),$$

donde $\nabla^2 f(\mathbf{x})$ es la matriz de $n \times n$ (llamada *matriz hessiana*) de la segunda derivada parcial de $f(\mathbf{x})$ evaluada en la solución de prueba actual **x** y $\left[\nabla^2 f(\mathbf{x})\right]^{-1}$ es la inversa de esta matriz hessiana.

Los algoritmos de programación no lineal que emplean el método de Newton —incluso aquellos que lo adaptan para ayudar a tratar con problemas de optimización *restringida*— por lo general aproximan el inverso de la función hessiana de varias maneras. Estas aproximaciones del método de Newton son conocidas como **métodos cuasi-Newton** (o *métodos métricos variables*). Se comentará a profundidad el papel importante de estos métodos en programación no lineal en la sección 12.9.

Una descripción más extensa de estos métodos está fuera de los alcances del libro, pero se pueden encontrar mayores detalles en textos dedicados a la programación no lineal.

12.6 CONDICIONES DE KARUSH-KUHN-TUCKER (KKT) PARA OPTIMIZACIÓN RESTRINGIDA

La pregunta ahora es cómo reconocer *una solución óptima* para un problema de programación no lineal (con funciones diferenciables). ¿Cuáles son las condiciones necesarias y (tal vez) suficientes que esa solución debe cumplir?

En las secciones anteriores se hicieron notar estas condiciones para *optimización no restringida*, como se resume en los primeros dos renglones de la tabla 12.4. Al principio de la sección 12.3 también se expusieron estas condiciones para la ligera *extensión* de optimización no restringida cuando *sólo* se tienen restricciones de no negatividad. En el tercer renglón de la tabla 12.4 se

TABLA 12.4 Condiciones necesarias y suficientes para la optimalidad

Problema	Condiciones necesarias para la optimalidad	También suficientes si:
Una variable, no restringido	$\frac{df}{dx} = 0$	$f(x)$ cóncava
Multivariables, no restringido	$\frac{\partial f}{\partial x_j} = 0 \quad (j = 1, 2, \ldots, n)$	$f(\mathbf{x})$ cóncava
Restringido, sólo restricciones de no negatividad	$\frac{\partial f}{\partial x_j} = 0 \quad (j = 1, 2, \ldots, n)$ (o ≤ 0 si $x_j = 0$)	$f(\mathbf{x})$ cóncava
Problema general restringido	Condiciones de Karush-Kuhn-Tucker	$f(\mathbf{x})$ cóncava y $g_i(\mathbf{x})$ convexa $(i = 1, 2, \ldots, m)$

muestran estas condiciones. Como se indica en el último renglón de la tabla, las condiciones para el caso general se llaman **condiciones de Karush-Kuhn-Tucker** (o **condiciones KKT**), porque fueron desarrolladas de manera independiente por Karush[11] y por Kuhn y Tucker.[12] Su resultado básico se expresa en el siguiente teorema.

Teorema. Suponga que $f(\mathbf{x}), g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_m(\mathbf{x})$ son funciones *diferenciables* que satisfacen ciertas condiciones de regularidad.[13] Entonces,

$$\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$$

puede ser una *solución óptima* para el problema de programación no lineal, sólo si existen m números $u_1, u_2, \ldots, u_m$, que satisfagan *todas* las siguientes *condiciones KKT*:

$$\left.\begin{array}{ll} \textbf{1.} & \dfrac{\partial f}{\partial x_j} - \displaystyle\sum_{i=1}^{m} u_i \dfrac{\partial g_i}{\partial x_j} \leq 0 \\ \textbf{2.} & x_j^* \left(\dfrac{\partial f}{\partial x_j} - \displaystyle\sum_{i=1}^{m} u_i \dfrac{\partial g_i}{\partial x_j} \right) = 0 \end{array}\right\} \text{ en } \mathbf{x} = \mathbf{x}^*, \text{ para } j = 1, 2, \ldots, n.$$

$$\left.\begin{array}{ll} \textbf{3.} & g_i(\mathbf{x}^*) - b_i \leq 0 \\ \textbf{4.} & u_i[g_i(\mathbf{x}^*) - b_i] = 0 \end{array}\right\} \quad \text{para } i = 1, 2, \ldots, m.$$

$$\textbf{5.} \quad x_j^* \geq 0, \qquad \text{para } j = 1, 2, \ldots, n.$$

$$\textbf{6.} \quad u_i \geq 0, \qquad \text{para } i = 1, 2, \ldots, m.$$

Observe que las condiciones 2 y 4 requieren que el producto de dos cantidades sea 0. Por tanto, cada una de estas condiciones en realidad dice que al menos una de las dos cantidades debe ser cero. En consecuencia, la condición 4 se puede combinar con la condición 3 para expresarlas en otra forma equivalente, tal como

$$(3, 4) \qquad g_i(\mathbf{x}^*) - b_i = 0$$
$$(\text{o} \leq 0 \quad \text{si } u_i = 0), \qquad \text{para } i = 1, 2, \ldots, m.$$

De manera similar, la condición 2 se puede combinar con la 1 de la siguiente manera:

$$(1, 2) \qquad \frac{\partial f}{\partial x_j} - \sum_{i=1}^{m} u_i \frac{\partial g_i}{\partial x_j} = 0$$
$$(\text{o} \leq 0 \quad \text{si } x_j^* = 0), \qquad \text{para } j = 1, 2, \ldots, n.$$

[11] W. Karush, "Minima of Functions of Several Variables with Inequalities as Side Conditions", M. S. thesis, Department of Mathematics, University of Chicago, 1939.

[12] H. W. Kuhn y A. W. Tucker: "Nonlinear Programming", en Jerzy Newman (ed.), *Proceedings of the Second Berkeley Symposium*, University of California Press, Berkeley, 1951, pp. 481-492.

[13] *Ibid.*, p. 483.

Cuando $m = 0$ (sin restricciones funcionales), esta suma se elimina y la condición combinada (1, 2) se reduce a la condición que se presenta en el tercer renglón de la tabla 12.4. Así, para $m > 0$, cada término de la suma modifica la condición de $m = 0$ para incorporar el efecto de la restricción funcional correspondiente.

En las condiciones 1, 2, 4 y 6, las u_i corresponden a las *variables duales* de programación lineal —al final de la sección se estudia más a fondo esta correspondencia— que tienen una interpretación económica comparable. En realidad, las u_i surgieron de la derivación matemática, como los *multiplicadores de Lagrange* (que se presentaron en el apéndice 3). Las condiciones 3 y 5 sólo ayudan a asegurar la factibilidad de la solución. Las otras condiciones eliminan la mayor parte de las soluciones factibles como posibles candidatos para una solución óptima.

Debe hacerse notar que el cumplimiento de estas condiciones no garantiza que la solución sea óptima. Como se resume en la última columna de la tabla 12.4, son necesarios ciertos supuestos de *convexidad* adicionales para obtener esta garantía. Estos supuestos se establecen de manera formal en la siguiente extensión del teorema.

Corolario. Suponga que $f(\mathbf{x})$ es una función *cóncava* y que $g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_m(\mathbf{x})$ son funciones *convexas* —es decir, se trata de un problema de programación convexa—, en donde todas estas funciones satisfacen las condiciones de regularidad. Entonces, $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$ es una *solución óptima* si y sólo si se satisfacen todas las condiciones del teorema.

Ejemplo. Para ilustrar esta formulación y la aplicación de las *condiciones KKT*, considere el siguiente problema de programación no lineal con dos variables:

$$\text{Maximizar} \qquad f(\mathbf{x}) = \ln(x_1 + 1) + x_2,$$

sujeta a

$$2x_1 + x_2 \leq 3$$

y

$$x_1 \geq 0, \quad x_2 \geq 0,$$

donde ln denota el logaritmo natural. De esta forma, $m = 1$ (una restricción funcional) y $g_1(\mathbf{x}) = 2x_1 + x_2$, de manera que $g_1(\mathbf{x})$ es convexa. Aún más, es fácil verificar (vea el apéndice 2) que $f(\mathbf{x})$ es cóncava. Entonces se puede aplicar el corolario, por lo que cualquier solución que satisfaga las condiciones KKT será en definitiva una solución óptima. Al aplicar las fórmulas en el teorema, se obtienen las siguientes condiciones KKT para este ejemplo:

1(j = 1). $\dfrac{1}{x_1 + 1} - 2u_1 \leq 0.$

2(j = 1). $x_1\left(\dfrac{1}{x_1 + 1} - 2u_1\right) = 0.$

1(j = 2). $1 - u_1 \leq 0.$

2(j = 2). $x_2(1 - u_1) = 0.$

3. $2x_1 + x_2 - 3 \leq 0.$

4. $u_1(2x_1 + x_2 - 3) = 0.$

5. $x_1 \geq 0, x_2 \geq 0.$

6. $u_1 \geq 0.$

A continuación se describen los pasos para resolver las condiciones KKT para este ejemplo.

1. $u_1 \geq 1$, de la condición 1(j = 2).
   $x_1 \geq 0$, por la condición 5.
2. Por lo tanto, $\dfrac{1}{x_1 + 1} - 2u_1 < 0.$
3. Por lo tanto, $x_1 = 0$, de la condición 2(j = 1).
4. $u_1 \neq 0$ implica que $2x_1 + x_2 - 3 = 0$, de la condición 4.
5. Los pasos 3 y 4 implican que $x_2 = 3$.

6. $x_2 \neq 0$ implica que $u_1 = 1$, de la condición 2(j = 2).
7. Los valores $x_1 = 0$, $x_2 = 3$ y $u_1 = 1$ no violan ninguna condición.

Por lo tanto, existe un número $u_1 = 1$ tal que $x_1 = 0$, $x_2 = 3$ y $u_1 = 1$ satisfacen todas las condiciones. En consecuencia, $\mathbf{x}^* = (0, 3)$ es una solución óptima para este problema.

Éste, en particular, fue un problema relativamente fácil de resolver porque los primeros dos pasos condujeron con rapidez a las conclusiones restantes. Con frecuencia resulta más difícil encontrar la forma de comenzar. La sucesión particular de pasos que se necesita para resolver las condiciones KKT difiere de un problema a otro. Cuando la lógica no es evidente, a veces es útil considerar por separado los distintos casos en donde cada x_j y u_i se especifican como iguales o mayores que 0, y después se prueba cada caso hasta que uno conduce a una solución.

Para ilustrar, suponga que este enfoque de considerar los diferentes casos de manera separada se ha aplicado al ejemplo previo en lugar de emplear la lógica implicada en los siete pasos anteriores. Para este ejemplo, hay ocho casos que corresponden a las ocho combinaciones de $x_1 = 0$ contra $x_1 > 0$, $x_2 = 0$ contra $x_2 > 0$, o $u_1 = 0$ contra $u_1 > 0$. Cada caso conduce a una afirmación y análisis más sencillos de las condiciones. Como ejemplo, considere primero el caso que se expone en seguida, en donde $x_1 = 0$, $x_2 = 0$ y $u_1 = 0$.

Condiciones KKT para el caso $x_1 = 0, x_2 = 0, u_1 = 0$

1(j = 1). $\frac{1}{0+1} \leq 0.$ Contradicción.
1(j = 2). $1 - 0 \leq 0.$ Contradicción.
3. $0 + 0 \leq 3.$

(Todas las demás condiciones son redundantes.)

Como se enumera después, los otros tres casos donde $u_1 = 0$ también llevan de inmediato a contradicciones en forma muy parecida, por lo cual no se dispone de soluciones.

Caso $x_1 = 0, x_2 > 0, u_1 = 0$ contradice las condiciones 1(j = 1), 1(j = 2) y 2(j = 2).
Caso $x_1 > 0, x_2 = 0, u_1 = 0$ contradice las condiciones 1(j = 1), 2(j = 1) y 1(j = 2).
Caso $x_1 > 0, x_2 > 0, u_1 = 0$ contradice las condiciones 1(j = 1), 2(j = 1), 1(j = 2) y 2(j = 2).

El caso $x_1 > 0, x_2 > 0, u_1 > 0$ permite eliminar estos tres multiplicadores diferentes de cero de las condiciones 2(j = 1), 2(j = 2) y 4, lo que a su vez permite eliminar las condiciones 1(j = 1), 1(j = 2) y 3 como redundantes, tal como se resume en seguida.

Condiciones KKT para el caso $x_1 > 0, x_2 > 0, u_1 > 0$

1(j = 1). $\frac{1}{x_1 + 1} - 2u_1 = 0.$
2(j = 2). $1 - u_1 = 0.$
4. $2x_1 + x_2 - 3 = 0.$
(Todas las demás condiciones son redundantes.)

Entonces, $u_1 = 1$, así $x_1 = -\frac{1}{2}$, lo que contradice $x_1 > 0$.

Ahora suponga que se intenta el caso $x_1 = 0, x_2 > 0, u_1 > 0$.

Condiciones KKT para el caso $x_1 = 0, x_2 > 0, u_1 > 0$

1(j = 1). $\frac{1}{0 + 1} - 2u_1 = 0.$
2(j = 2). $1 - u_1 = 0.$
4. $0 + x_2 - 3 = 0.$

(Todas las demás condiciones son redundantes.)

Por lo tanto, $x_1 = 0$, $x_2 = 3$, $u_1 = 1$. Una vez que se ha encontrado una solución, se sabe que no es necesario considerar casos adicionales.

Si desea ver **otro ejemplo** de las condiciones KKT para encontrar una solución óptima, en la sección Worked Examples del CD-ROM se proporciona uno más.

Para problemas más complicados que éste puede ser difícil, si no es que materialmente imposible, derivar una solución óptima *directa* de las condiciones KKT. De todas maneras, estas condiciones proporcionan información valiosa en cuanto a la identidad de una solución óptima y también permiten verificar que una solución propuesta pueda ser óptima.

Existen muchas aplicaciones *indirectas* valiosas de las condiciones KKT. Una de ellas surge de la *teoría de dualidad* que se desarrolló para programación no lineal, que se puede poner en paralelo con la teoría de dualidad para programación lineal que se presentó en el capítulo 6. En particular para cualquier problema de maximización restringida (llamado *problema primal*), se pueden usar las condiciones KKT para definir un problema dual asociado, que es un problema de minimización restringido. Las variables del problema dual consisten en multiplicadores de Lagrange, u_i ($i = 1, 2, \ldots, m$) y las variables primales x_j ($j = 1, 2, \ldots, n$). En el caso especial de que el problema primal es un problema de programación lineal, las variables x_j desaparecerán del problema dual y quedará el ya conocido problema dual de programación lineal, en el que las variables u_i corresponden a las variables y_i del capítulo 6. Cuando un problema primal es un problema de programación convexa, es posible establecer relaciones entre el problema primal y el problema dual, que son muy parecidas a las de programación lineal. Por ejemplo, la *propiedad de dualidad fuerte* de la sección 6.1, que establece que los valores óptimos de las funciones objetivo son iguales, también se cumple aquí. Más aún, los valores de las variables u_i en una solución óptima para el problema dual, de nuevo se pueden interpretar como los *precios sombra* (vea las secciones 4.7 y 6.2); es decir, proporcionan la tasa a la que se puede aumentar el valor óptimo de la función objetivo del problema primal, si se aumenta (ligeramente) el lado derecho de la restricción correspondiente. Como la teoría de dualidad para programación no lineal es un tema relativamente avanzado, se pide al lector interesado que consulte otros libros para obtener más información.[14]

En la sección siguiente se verá otra aplicación indirecta de las condiciones KKT.

12.7 PROGRAMACIÓN CUADRÁTICA

Como se indicó en la sección 12.3, el problema de programación cuadrática difiere del problema de programación lineal nada más en que la función objetivo incluye también términos x_j^2 y x_ix_j ($i \neq j$). En consecuencia, si se usa la notación matricial como la que se introdujo al inicio de la sección 5.2, el problema es encontrar $\mathbf{x}$ para

$$\text{Maximizar} \qquad f(\mathbf{x}) = \mathbf{cx} - \frac{1}{2}\mathbf{x}^T\mathbf{Qx},$$

sujeta a

$$\mathbf{Ax} \leq \mathbf{b} \qquad \text{y} \qquad \mathbf{x} \geq \mathbf{0},$$

donde $\mathbf{c}$ es un vector renglón y $\mathbf{x}$ y $\mathbf{b}$ son vectores columna. $\mathbf{Q}$ y $\mathbf{A}$ son matrices y el superíndice T denota la transpuesta (vea el apéndice 4). Las q_{ij} (elementos de Q) son constantes dadas tales que $q_{ij} = q_{ji}$ (que es la razón para que aparezca el factor de $\frac{1}{2}$ en la función objetivo). Al realizar las multiplicaciones entre vectores y matrices indicadas, la función objetivo se expresa en términos de estas q_{ij}, las c_j (elementos de $\mathbf{c}$) y las variables, de la siguiente manera:

$$f(\mathbf{x}) = \mathbf{cx} - \frac{1}{2}\mathbf{x}^T\mathbf{Qx} = \sum_{j=1}^{n} c_jx_j - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij}x_ix_j.$$

Para cada término donde $i = j$ en esta doble suma, $x_ix_j = x_j^2$, así $-\frac{1}{2}q_{jj}$ es el coeficiente de x_j^2. Cuando $i \neq j$, entonces $-\frac{1}{2}(q_{ij}x_ix_j + q_{ji}x_jx_i) = -q_{ij}x_ix_j$, de manera que $-q_{ij}$ es el coeficiente total del producto de x_i y x_j.

Para ilustrar esta notación, considere el siguiente ejemplo de un problema de programación cuadrática.

[14] Una versión unificada de varios enfoques sobre la dualidad en programación no lineal se encuentra en A. M. Geoffrion, "Duality in Nonlinear Programming: A Simplified Applications-Oriented Development", en *SIAM Review*, **13**: 1-37, 1971.

Maximizar $\quad f(x_1, x_2) = 15x_1 + 30x_2 + 4x_1x_2 - 2x_1^2 - 4x_2^2,$

sujeta a

$$x_1 + 2x_2 \leq 30$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

En este caso,

$$\mathbf{c} = [15 \quad 30], \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad \mathbf{Q} = \begin{bmatrix} 4 & -4 \\ -4 & 8 \end{bmatrix},$$

$$\mathbf{A} = [1 \quad 2], \qquad \mathbf{b} = [30].$$

Observe que

$$\begin{aligned}
\mathbf{x}^T\mathbf{Q}\mathbf{x} &= [x_1 \quad x_2] \begin{bmatrix} 4 & -4 \\ -4 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= [(4x_1 - 4x_2) \quad (-4x_1 + 8x_2)] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= 4x_1^2 - 4x_2x_1 - 4x_1x_2 + 8x_2^2 \\
&= q_{11}x_1^2 + q_{21}x_2x_1 + q_{12}x_1x_2 + q_{22}x_2^2.
\end{aligned}$$

Al multiplicar por $-\frac{1}{2}$ se obtiene

$$-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} = -2x_1^2 + 4x_1x_2 - 4x_2^2,$$

que es la porción no lineal de la función objetivo de este ejemplo. Como $q_{11} = 4$ y $q_{22} = 8$, el ejemplo muestra que $-\frac{1}{2}q_{jj}$ es el coeficiente de x_j^2 de la función objetivo. El hecho de que $q_{12} = q_{21} = -4$ ilustra que (ambos) $-q_{ij}$ y $-q_{ji}$ proporcionan el coeficiente total del producto de x_i y x_j.

Se dispone de varios algoritmos para el caso especial de un problema de programación cuadrática donde la función objetivo es una función *cóncava*. (Una manera de comprobar que la función objetivo sea cóncava es verificar la condición equivalente de que

$$\mathbf{x}^T\mathbf{Q}\mathbf{x} \geq 0$$

para toda **x**, es decir, que **Q** es una matriz *positiva semidefinida*.) Se describirá uno[15] de estos algoritmos, el *método símplex modificado*, que es muy aceptado porque se basa en el método símplex con una pequeña modificación. La clave de este enfoque es construir las condiciones KKT de la sección anterior y después reexpresarlas de una manera conveniente que se parezca a programación lineal. Por esto, antes de describir el algoritmo, se desarrollará esta práctica forma.

Condiciones KKT para programación cuadrática

Para concretar, primero considere el ejemplo anterior. A partir de la forma que se presentó en la sección anterior, sus condiciones KKT son las siguientes.

1(j = 1). $15 + 4x_2 - 4x_1 - u_1 \leq 0.$
2(j = 1). $x_1(15 + 4x_2 - 4x_1 - u_1) = 0.$
1(j = 2). $30 + 4x_1 - 8x_2 - 2u_1 \leq 0.$
2(j = 2). $x_2(30 + 4x_1 - 8x_2 - 2u_1) = 0.$
3. $x_1 + 2x_2 - 30 \leq 0.$
4. $u_1(x_1 + 2x_2 - 30) = 0.$
5. $x_1 \geq 0, \qquad x_2 \geq 0.$
6. $u_1 \geq 0.$

[15] P. Wolfe, "The Simplex Method for Quadratic Program ming", en *Econometrics*, **27**: 382-398, 1959. Este artículo desarrolla tanto una forma corta como una larga del algoritmo. Aquí se presenta una versión de la *forma corta* que tiene el supuesto adicional de que o bien $\mathbf{c} = \mathbf{0}$ *o bien* la función objetivo es estrictamente cóncava.

Para poder reexpresar estas condiciones en una forma más conveniente, se mueven las constantes de las condiciones 1(j = 1), 1(j = 2) y 3 al lado derecho y después se introducen *variables de holgura* no negativas (denotadas por y_1, y_2 y v_1, respectivamente) para convertir estas desigualdades en ecuaciones.

$$\begin{array}{ll} \mathbf{1(j=1).} & -4x_1 + 4x_2 - u_1 + y_1 = -15 \\ \mathbf{1(j=2).} & 4x_1 - 8x_2 - 2u_1 + y_2 = -30 \\ \mathbf{3.} & x_1 + 2x_2 + v_1 = 30 \end{array}$$

Observe que la condición 2(j = 1) se puede reexpresar ahora en forma simple como la necesidad de que se cumpla una de las dos, $x_1 = 0$ o $y_1 = 0$; esto es,

$$\mathbf{2(j=1).}\ x_1y_1 = 0.$$

Exactamente en la misma forma, las condiciones 2(j = 2) y 4 se pueden sustituir por

$$\begin{array}{ll} \mathbf{2(j=2).} & x_2y_2 = 0, \\ \mathbf{4.} & u_1v_1 = 0. \end{array}$$

Para cada uno de estos tres pares —(x_1, y_1), (x_2, y_2), (u_1, v_1)— las dos variables se llaman **variables complementarias**, porque sólo una de las dos puede ser diferente de cero. Estas nuevas formas de las condiciones 2(j = 1), 2(j = 2) y 4 se pueden combinar en una restricción,

$$x_1y_1 + x_2y_2 + u_1v_1 = 0,$$

llamada **restricción de complementariedad**.

Después de multiplicar las ecuaciones de las condiciones 1(j = 1) y 1(j = 2) por –1 para obtener lados derechos no negativos, se obtiene la forma deseada del conjunto completo de condiciones:

$$\begin{array}{l} 4x_1 - 4x_2 + u_1 - y_1 = 15 \\ -4x_1 + 8x_2 + 2u_1 - y_2 = 30 \\ x_1 + 2x_2 + v_1 = 30 \\ x_1 \geq 0, \quad x_2 \geq 0, \quad u_1 \geq 0, \quad y_1 \geq 0, \quad y_2 \geq 0, \quad v_1 \geq 0 \\ x_1y_1 + x_2y_2 + u_1v_1 = 0 \end{array}$$

Esta forma es en especial conveniente puesto que excepto por la restricción de complementariedad, estas *condiciones son restricciones de programación lineal*.

En *cualquier* problema de programación cuadrática, sus condiciones KKT se pueden reducir a la misma forma conveniente que contiene sólo restricciones de programación lineal y una restricción de complementariedad. En notación matricial, esta forma general es

$$\begin{array}{l} \mathbf{Qx} + \mathbf{A}^T\mathbf{u} - \mathbf{y} = \mathbf{c}^T, \\ \mathbf{Ax} + \mathbf{v} = \mathbf{b}, \\ \mathbf{x} \geq \mathbf{0}, \quad \mathbf{u} \geq \mathbf{0}, \quad \mathbf{y} \geq \mathbf{0}, \quad \mathbf{v} \geq \mathbf{0}, \\ \mathbf{x}^T\mathbf{y} + \mathbf{u}^T\mathbf{v} = 0, \end{array}$$

donde los elementos del vector columna **u** son las u_i de la sección anterior y los elementos de los vectores columna **y** y **v** son variables de holgura.

Como se supone que la función objetivo del problema original es cóncava y como las funciones de restricción son lineales y, por tanto, convexas, se puede aplicar el corolario del teorema de la sección 12.6. Así, **x** es *óptima* si y sólo si existen valores de **y**, **u** y **v** tales que los cuatro vectores juntos satisfacen todas estas condiciones. Así pues, el problema original se reduce al problema equivalente de encontrar una *solución factible* para estas *restricciones*.

Es interesante observar que este problema equivalente es un ejemplo del problema de *complementariedad lineal* que se introdujo en la sección 12.3 (vea el problema 12.3-6), y que una restricción clave de este problema de complementariedad lineal es su *restricción de complementariedad*.

Método símplex modificado

El *método símplex modificado* explota el importante hecho de que, con excepción de la restricción de complementariedad, las condiciones KKT expresadas en la forma conveniente que se acaba de

obtener, no son otra cosa que restricciones de programación lineal. Aún más, la restricción de complementariedad simplemente implica que no es permisible que las *dos* variables complementarias de un par dado sean básicas —no degeneradas, las únicas variables > 0— al considerar soluciones BF (no degeneradas). Por tanto, el problema se reduce a encontrar una solución inicial BF para cualquier problema de programación lineal con estas restricciones, sujeta a la restricción adicional sobre la identidad de las variables básicas. (Esta solución inicial básica factible puede ser la única solución factible en este caso.)

Como se analizó en la sección 4.6, tal solución BF inicial se encuentra casi de modo directo. En el sencillo caso de que $\mathbf{c}^T \leq \mathbf{0}$ (poco probable) y $\mathbf{b} \geq \mathbf{0}$, las variables básicas iniciales son los elementos de $\mathbf{y}$ y $\mathbf{v}$ —se multiplica por –1 todo el primer conjunto de ecuaciones—, de manera que la solución deseada es $\mathbf{x} = \mathbf{0}$, $\mathbf{u} = \mathbf{0}$, $\mathbf{y} = -\mathbf{c}^T$, $\mathbf{v} = \mathbf{b}$. De otra manera, es necesario revisar el problema e introducir una *variable artificial* en cada una de las ecuaciones donde $c_j > 0$ —se agrega la variable de la izquierda— o $b_i < 0$ —se resta la variable de la izquierda y después se multiplican los dos lados por –1—, con el fin de usar estas variables artificiales —que se denotan por z_1, z_2, etc.— como variables básicas iniciales del problema revisado. (Observe que esta selección de variables básicas iniciales satisface la restricción de complementariedad, puesto que como variables no básicas $\mathbf{x} = \mathbf{0}$ y $\mathbf{u} = \mathbf{0}$, de manera automática.)

Después se utiliza la fase 1 del *método de dos fases* (vea la sección 4.6) para encontrar una solución BF para el problema real; es decir, se aplica el método símplex (con una modificación) al siguiente problema de programación lineal

$$\text{Minimizar} \qquad Z = \sum_j z_j,$$

sujeta a las restricciones anteriores de programación lineal que se obtuvieron de las condiciones KKT, pero con estas variables artificiales incluidas.

La modificación del método símplex es el siguiente cambio en el procedimiento para elegir una variable básica entrante.

> **Regla de entrada restringida:** cuando se elige la variable básica entrante, se excluye de las posibilidades cualquier variable no básica cuya *variable complementaria* sea básica; la elección debe hacerse entre las otras variables no básicas, según el criterio normal del método símplex.

Esta regla conserva satisfecha la restricción de complementariedad en todo el curso del algoritmo. Cuando se obtiene una solución óptima

$$\mathbf{x}^*, \mathbf{u}^*, \mathbf{y}^*, \mathbf{v}^*, z_1 = 0, \ldots, z_n = 0$$

para la fase 1 de este problema, $\mathbf{x}^*$ es la solución óptima deseada para el problema original de programación cuadrática. La fase 2 del método de dos fases no se necesita.

Ejemplo. Se ilustrará este enfoque con el problema que se presentó al principio de la sección. Como se puede verificar con los resultados del apéndice 2 [vea el problema 12.7-1*a*], $f(x_1, x_2)$ es *estrictamente cóncava*, es decir,

$$\mathbf{Q} = \begin{bmatrix} 4 & -4 \\ -4 & 8 \end{bmatrix}$$

es positiva definida, por lo que se puede aplicar el algoritmo.

El punto de partida para resolver este ejemplo son las condiciones KKT en la forma conveniente que se obtuvo con anterioridad. Una vez que se introducen las variables artificiales necesarias, el problema de programación lineal que se resolverá en forma explícita por el método símplex modificado es

$$\text{Minimizar} \quad Z = z_1 + z_2,$$

sujeta a

$$\begin{aligned}
4x_1 - 4x_2 + u_1 - y_1 \qquad\qquad + z_1 \qquad &= 15 \\
-4x_1 + 8x_2 + 2u_1 \qquad - y_2 \qquad\qquad + z_2 &= 30 \\
x_1 + 2x_2 \qquad\qquad\qquad + v_1 \qquad\qquad &= 30
\end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad u_1 \geq 0, \quad y_1 \geq 0, \quad y_2 \geq 0, \quad v_1 \geq 0,$$
$$z_1 \geq 0, \quad z_2 \geq 0.$$

La restricción de complementariedad adicional

$$x_1y_1 + x_2y_2 + u_1v_1 = 0,$$

no se incluye de manera explícita, porque el algoritmo refuerza esta restricción de manera automática con la *regla de entrada restringida*. En particular, en cada uno de los tres pares de variables complementarias —(x_1, y_1), (x_2, y_2), (u_1, v_1)— siempre que una de las dos variables sea ya una variable básica, la otra se excluye de las candidatas como variable básica entrante. Recuerde que sólo las variables *diferentes de cero* son variables básicas. Dado que el conjunto inicial de variables básicas del problema de programación lineal —z_1, z_2, v_1— proporciona una solución BF inicial que satisface la restricción de complementariedad, no hay manera de violar esta restricción con ninguna de las soluciones BF subsecuentes.

En la tabla 12.5 se muestran los resultados de la aplicación del método símplex modificado a este problema. En la primera tabla símplex se presenta el sistema de ecuaciones inicial *después* de convertir de minimización de Z a maximización de $-Z$ y de eliminar algebraicamente las variables básicas iniciales de la ecuación (0), lo mismo que se hizo en el ejemplo de terapia de radiación en la sección 4.6. Las tres iteraciones proceden igual que en el método símplex normal, *excepto* por la eliminación de ciertos candidatos para la variable básica entrante, según la regla de entrada restringida. En la primera tabla se elimina u_1 como candidato porque su variable complementaria (v_1) es ya una variable básica (de todas maneras, habría entrado x_2 puesto que $-4 < -3$). En la segunda

■ **TABLA 12.5 Aplicación del método símplex modificado al ejemplo de programación cuadrática**

Iteración	Variable básica	Ec.	Z	x_1	x_2	u_1	y_1	y_2	v_1	z_1	z_2	Lado derecho
0	Z	(0)	−1	0	−4	−3	1	1	0	0	0	−45
	z_1	(1)	0	4	−4	1	−1	0	0	1	0	15
	z_2	(2)	0	−4	8	2	0	−1	0	0	1	30
	v_1	(3)	0	1	2	0	0	0	1	0	0	30
1	Z	(0)	−1	−2	0	−2	1	$\frac{1}{2}$	0	0	$\frac{1}{2}$	−30
	z_1	(1)	0	2	0	2	−1	$-\frac{1}{2}$	0	1	$\frac{1}{2}$	30
	x_2	(2)	0	$-\frac{1}{2}$	1	$\frac{1}{4}$	0	$-\frac{1}{8}$	0	0	$\frac{1}{8}$	$3\frac{3}{4}$
	v_1	(3)	0	2	0	$-\frac{1}{2}$	0	$\frac{1}{4}$	1	0	$-\frac{1}{4}$	$22\frac{1}{2}$
2	Z	(0)	−1	0	0	$-\frac{5}{2}$	1	$\frac{3}{4}$	1	0	$\frac{1}{4}$	$-7\frac{1}{2}$
	z_1	(1)	0	0	0	$\frac{5}{2}$	−1	$-\frac{3}{4}$	−1	1	$\frac{3}{4}$	$7\frac{1}{2}$
	x_2	(2)	0	0	1	$\frac{1}{8}$	0	$-\frac{1}{16}$	$\frac{1}{4}$	0	$\frac{1}{16}$	$9\frac{3}{8}$
	x_1	(3)	0	1	0	$-\frac{1}{4}$	0	$\frac{1}{8}$	$\frac{1}{2}$	0	$-\frac{1}{8}$	$11\frac{1}{4}$
3	Z	(0)	−1	0	0	0	0	0	0	1	1	0
	u_1	(1)	0	0	0	1	$-\frac{2}{5}$	$-\frac{3}{10}$	$-\frac{2}{5}$	$\frac{2}{5}$	$\frac{3}{10}$	3
	x_2	(2)	0	0	1	0	$\frac{1}{20}$	$-\frac{1}{40}$	$\frac{3}{10}$	$-\frac{1}{20}$	$\frac{1}{40}$	9
	x_1	(3)	0	1	0	0	$-\frac{1}{10}$	$\frac{1}{20}$	$\frac{2}{5}$	$\frac{1}{10}$	$-\frac{1}{20}$	12

tabla se elimina tanto u_1 *como* y_2 —puesto que v_1 y x_2 son variables básicas— y de manera automática se elige x_1 como único candidato con coeficientes negativos en el renglón 0, mientras que el método símplex *normal* habría permitido la entrada de x_1 *o bien* de u_1, porque están empatadas con el coeficiente más negativo. En la tercera tabla se eliminan y_1 y y_2 (porque x_1 y x_2 son variables básicas). Esta vez u_1 no se elimina, puesto que v_1 ya no es una variable básica, así que se elige u_1 como variable entrante.

La solución óptima que resulta para esta fase 1 del problema es $x_1 = 12$, $x_2 = 9$, $u_1 = 3$, mientras el resto de las variables es igual a cero. (El problema 12.7-l*c* pide que se verifique que esta solución es óptima por medio de demostrar que si $x_1 = 12$, $x_2 = 9$ y $u_1 = 3$, se satisfacen las condiciones KKT del problema original, cuando se escriben en la forma dada en la sección 12.6.) Entonces, la solución óptima para el problema de programación cuadrática —que sólo incluye las variables x_1 y x_2— es $(x_1, x_2) = (12, 9)$.

En la sección Worked Examples de la página en internet de este libro se incluye **otro ejemplo** que ilustra la aplicación del método símplex modificado a un problema de programación cuadrática. En este ejemplo también se aplican las condiciones KKT.

Algunas opciones de software

En el IOR Tutorial se incluye una rutina interactiva del método símplex modificado para ayudar a aprender este algoritmo. Además, tanto Excel como LINGO, LINDO y MPL/CPLEX pueden resolver problemas de programación cuadrática.

El procedimiento para usar Excel es casi el mismo que se emplea en programación lineal. La única diferencia crucial es que la ecuación que se introduce a la celda que contiene el valor de la función objetivo ahora debe ser una ecuación cuadrática. Para ilustrar este caso, considere de nuevo el ejemplo que se introdujo al principio de la sección, que tiene la función objetivo

$$f(x_1, x_2) = 15x_1 + 30x_2 + 4x_1x_2 - 2x_1^2 - 4x_2^2.$$

Suponga que los valores de x_1 y x_2 están en las celdas B4 y C4 de la hoja de cálculo de Excel, y que el valor de la función objetivo está en la celda F4. En este caso, la ecuación para la celda F4 debe ser

F4 = 15*B4 + 30*C4 + 4*B4*C4 – 2*(B4^2) – 4*(C4^2),

donde el símbolo ^2 indica un exponente de 2. Antes de resolver el modelo, se debe oprimir el botón opciones y asegurarse de que la opción *Assume Linear Model* (Adoptar modelo lineal) *no* esté seleccionada (puesto que éste no es un modelo de programación *lineal*).

Cuando se use MPL/CPLEX se debe establecer el tipo de modelo como cuadrático mediante la introducción del siguiente enunciado al inicio del archivo del modelo:

OPTIONS

ModelType = Quadratic

(En forma alternativa, se puede seleccionar la opción Quadratic Models en el cuadro de diálogo de MPL Language, pero debe cambiar la configuración cuando de nuevo se trabaje con un modelo lineal.) Por lo demás, el procedimiento es el mismo que para programación lineal, excepto que la expresión de la función objetivo ahora es una función cuadrática. Así, en el ejemplo, la función objetivo se expresaría como

15x1 + 30x2 + 4x1*x2 – 2(x1^2) – 4(x2^2).

No es necesario hacer más al correr CPLEX, pues reconocerá de manera automática que el modelo se refiere a un problema de programación cuadrática.

Esta función objetivo se debe expresar de la misma manera que para un modelo de LINGO. Después LINGO hará una llamada automática a su solucionador no lineal para resolverlo.

En realidad, todos los archivos de Excel, MPL/CPLEX y LINGO/LINDO para este capítulo del OR Courseware demuestran sus procedimientos, pues exhiben los detalles de la forma en que estos paquetes de software establecen y resuelven el ejemplo.

Algunos de estos paquetes se pueden aplicar a tipos más complicados de problemas de programación no lineal que la programación cuadrática. Aunque CPLEX no puede, la versión profesional de MPL tiene otros solucionadores que sí lo hacen. La versión para estudiantes de MPL que se encuentra en la página en internet de este libro, incluye un solucionador llamado CONOPT (un producto de ARKI Consulting) que está diseñado para resolver problemas de programación convexa. Se puede emplear después de agregar el siguiente enunciado al inicio del archivo del modelo.

OPTIONS

ModelType = Nonlinear

Tanto Excel como LINGO incluyen solucionadores no lineales versátiles. Sin embargo, debe tenerse en mente que no se garantiza la obtención de una solución óptima para problemas complicados, en especial para problemas de programación no convexa (el tema de la sección 12.10). Por otro lado, LINGO contiene un *optimizador global* que encontrará una solución óptima global para problemas de programación no convexa que sean lo suficientemente pequeños. MPL también contiene un optimizador global llamado LGO como uno de los solucionadores que se proporcionan en el sitio en internet de este libro.

12.8 PROGRAMACIÓN SEPARABLE

En la sección anterior se presentó la forma en que se puede resolver una clase de problemas de programación no lineal mediante una extensión del método símplex. Ahora se estudiará otra clase llamada *programación separable*, para la que se puede encontrar una aproximación tan cercana como se quiera con un problema de programación lineal que tiene un número mayor de variables.

Como se indicó en la sección 12.3, la programación separable supone que la función objetivo $f(\mathbf{x})$ es cóncava, todas las restricciones $g_i(\mathbf{x})$ son convexas y todas estas funciones son funciones separables (lo que significa que cada término incluye una sola variable). Sin embargo, para simplificar la presentación, esta sección se centra en el caso especial en el que las funciones $g_i(\mathbf{x})$ convexas y separables, en realidad son *funciones lineales* como las de programación lineal. (Al final de esta sección volveremos a revisar el caso general de manera breve.) En consecuencia, sólo la función objetivo requerirá un tratamiento especial.

Bajo los supuestos anteriores, la función objetivo se puede expresar como una suma de funciones cóncavas de variables individuales,

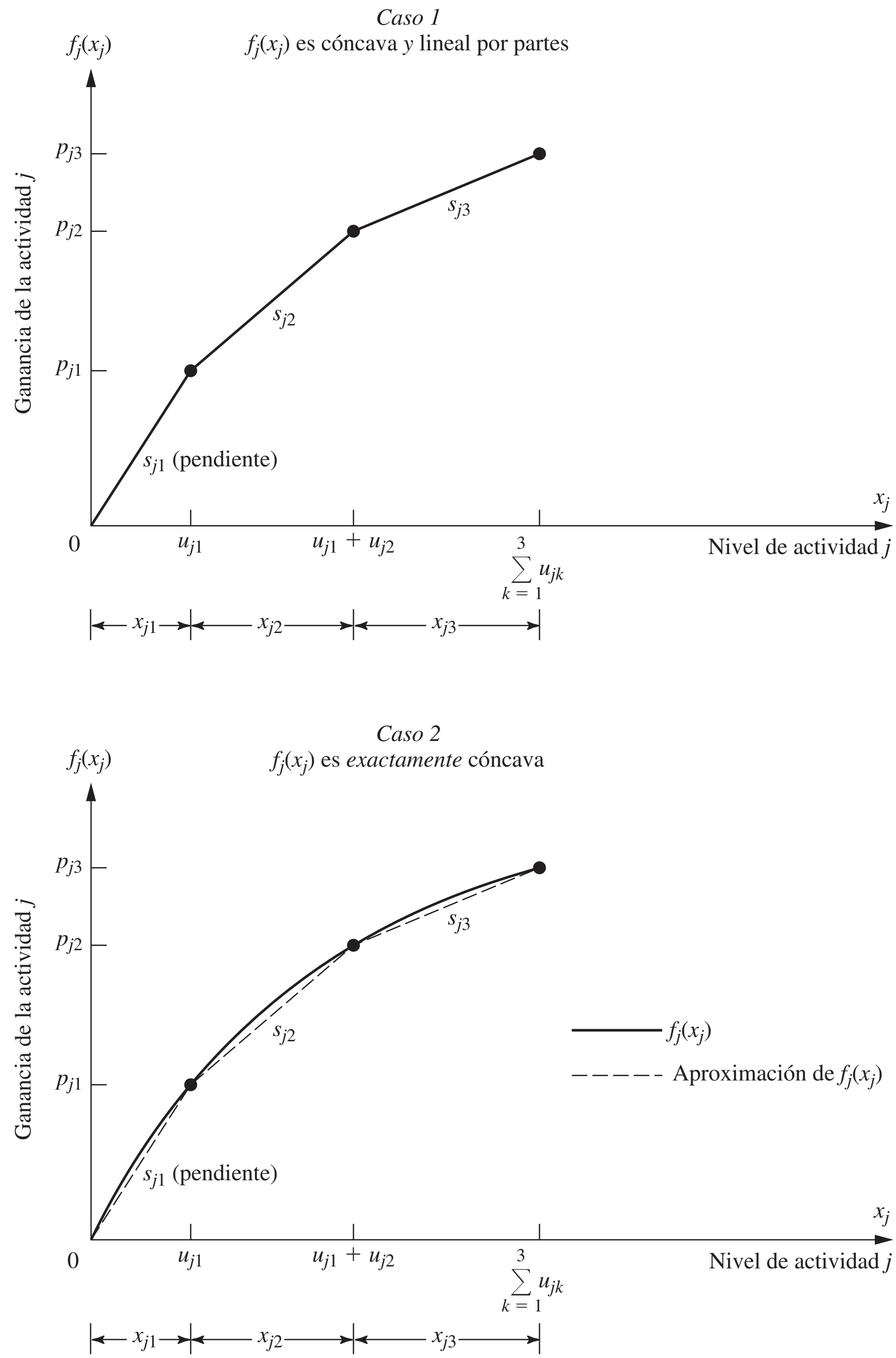
$$f(\mathbf{x}) = \sum_{j=1}^{n} f_j(x_j),$$

de manera que cada $f_j(x_j)$ tiene una forma[16] como la que se muestra en la figura 12.15, en el rango de valores factibles de x_j. Como $f(\mathbf{x})$ representa la medida de desempeño (por ejemplo, ganancia) de todas las actividades juntas, $f_j(x_j)$ representa la *contribución a la ganancia* por parte de la actividad j cuando se realiza al nivel x_j. La condición de que $f(\mathbf{x})$ sea separable sólo implica aditividad (vea la sección 3.3); esto es, no existe interacción entre las actividades —no existen términos de productos cruzados— que afecten la ganancia total más allá de sus contribuciones independientes. El supuesto de que cada $f_j(x_j)$ es cóncava indica que la *ganancia marginal* —pendiente de la curva de ganancia—, se mantiene igual o decrece (*nunca* aumenta) a medida que x_j crece.

Las curvas cóncavas de la ganancia se presentan con mucha frecuencia. Por ejemplo, puede ser posible vender una cantidad limitada de algún producto a un cierto precio y después una cantidad adicional a un precio menor, y tal vez, otra cantidad adicional a un precio todavía menor. De igual manera, puede ser necesario comprar materias primas a fuentes cada vez más costosas. Otra situación común es aquella en la que debe emplearse un proceso de producción más caro —como el uso de tiempo extra—, para aumentar la tasa de producción sobre cierta cantidad.

Este tipo de situaciones puede conducir a cualquiera de las curvas de ganancia que se muestran en la figura 12.15. En el caso 1, la pendiente decrece *por intervalos*, de manera que $f_j(x_j)$ es una *función lineal por partes* (una sucesión de segmentos de recta conectados). En el caso 2, la

[16] $f(\mathbf{x})$ es cóncava si y sólo si *cada* $f_j(x_j)$ es cóncava.



■ FIGURA 12.15
Forma de curvas de ganancia para programación separable.

pendiente puede decrecer en forma continua al aumentar x_j, de manera que $f_j(x_j)$ es una función cóncava general. Cualquier función de este tipo se puede aproximar tanto como se quiera por una función lineal por partes y este tipo de aproximación se usa, según sea necesario, para problemas de programación separable. (En la figura 12.15 se muestra una función de aproximación que consiste en sólo tres segmentos de recta, pero la aproximación se puede mejorar si se introducen más intervalos.) Esta aproximación es muy conveniente, puesto que una función lineal por partes de una sola variable se puede escribir como una *función lineal* de múltiples variables, con la restricción especial sobre estas variables que se describe a continuación.

Reformulación como un problema de programación lineal

La clave para dar a una función lineal por partes la forma de una función lineal es usar variables separadas para cada segmento de recta. Como ilustración, considere la función lineal por partes $f_j(x_j)$ que se muestra en la figura 12.15, caso 1 —o la aproximación por la función lineal por partes del caso 2—, que tiene tres segmentos de recta en el intervalo de valores factibles de x_j. Se introducen las tres nuevas variables, x_{j1}, x_{j2} y x_{j3}, y se establece

$$x_j = x_{j1} + x_{j2} + x_{j3},$$

donde

$$0 \leq x_{j1} \leq u_{j1}, \qquad 0 \leq x_{j2} \leq u_{j2}, \qquad 0 \leq x_{j3} \leq u_{j3}.$$

Después se usan las pendientes, s_{j1}, s_{j2} y s_{j3} para volver a escribir $f_j(x_j)$ como

$$f_j(x_j) = s_{j1}x_{j1} + s_{j2}x_{j2} + s_{j3}x_{j3},$$

Con la **restricción especial**

$$\begin{aligned} x_{j2} &= 0 \qquad \text{siempre que} \qquad x_{j1} < u_{j1}, \\ x_{j3} &= 0 \qquad \text{siempre que} \qquad x_{j2} < u_{j2}. \end{aligned}$$

Para entender por qué se requiere esta restricción especial, suponga que $x_j = 1$, donde $u_{jk} > 1 (k = 1, 2, 3)$, de manera que $f_j(1) = s_{j1}$. Observe que

$$x_{j1} + x_{j2} + x_{j3} = 1$$

permite

$$\begin{aligned} x_{j1} &= 1, \quad x_{j2} = 0, \quad x_{j3} = 0 \quad \Rightarrow \quad f_j(1) = s_{j1}, \\ x_{j1} &= 0, \quad x_{j2} = 1, \quad x_{j3} = 0 \quad \Rightarrow \quad f_j(1) = s_{j2}, \\ x_{j1} &= 0, \quad x_{j2} = 0, \quad x_{j3} = 1 \quad \Rightarrow \quad f_j(1) = s_{j3}, \end{aligned}$$

y así sucesivamente, donde

$$s_{j1} > s_{j2} > s_{j3}.$$

Sin embargo, la restricción especial sólo permite la primera posibilidad, que es la única que da el valor correcto de $f_j(1)$.

Desafortunadamente, la restricción especial no se ajusta al formato que requieren las restricciones de programación lineal, por lo cual *algunas* funciones lineales por partes no se pueden reescribir en el formato de programación lineal. Sin embargo, se supone que *nuestra* $f_j(x_j)$ es cóncava, por lo que $s_{j1} > s_{j2} > \ldots$, por lo cual un algoritmo para maximizar $f(\mathbf{x})$ de manera *automática* asigna la prioridad más alta al uso de x_{j1} cuando (en realidad) aumenta el valor de x_j desde cero, la siguiente prioridad al uso de x_{j2}, etc., sin incluir siquiera la restricción especial explícita en el modelo. Esta observación conduce a la siguiente propiedad esencial.

Propiedad esencial de programación separable. Cuando $f(\mathbf{x})$ y $g_i(\mathbf{x})$ satisfacen los supuestos de programación separable y las funciones lineales por partes que resultan se ponen en la forma de funciones lineales, al eliminar la *restricción especial* se obtiene un *modelo de programación lineal* cuya solución óptima satisface de manera automática la restricción especial.

Más adelante, en esta sección, se profundizará un poco más en la lógica que sustenta esta propiedad esencial, en el contexto de un ejemplo específico. (Vea también el problema 12.8-6*a*.)

Para escribir el modelo completo de programación lineal con la notación anterior, sea n_j el número de segmentos de recta en $f_j(x_j)$ (o en la función lineal por partes que la aproxima), de manera que

$$x_j = \sum_{k=1}^{n_j} x_{jk}$$

se sustituirá en todo el modelo original y

$$f_j(x_j) = \sum_{k=1}^{n_j} s_{jk} x_{jk}$$

se sustituirá[17] en la función objetivo para $j = 1, 2, \ldots, n$. El modelo que se obtiene es

$$\text{Maximizar} \qquad Z = \sum_{j=1}^{n} \left(\sum_{k=1}^{n_j} s_{jk} x_{jk} \right),$$

sujeta a

$$\sum_{j=1}^{n} a_{ij} \left(\sum_{k=1}^{n_j} x_{jk} \right) \leq b_i, \qquad \text{para } i = 1, 2, \ldots, m$$

$$x_{jk} \leq u_{jk}, \qquad \text{para } k = 1, 2, \ldots, n_j;\ j = 1, 2, \ldots, n$$

y

$$x_{jk} \geq 0, \qquad \text{para} \qquad k = 1, 2, \ldots, n_j;\ j = 1, 2, \ldots, n.$$

(Las restricciones $\Sigma_{k=1}^{n_j} x_{jk} \geq 0$ se eliminan, puesto que están garantizadas por las restricciones $x_{jk} \geq 0$.) Si algunas variables originales x_j no tuvieran cota superior, entonces $u_{jn_j} = \infty$, por lo que se debe eliminar la restricción que incluye esta cantidad.

Una manera eficiente de resolver este modelo[18] es usar la versión simplificada del método símplex para manejar restricciones de cota superior (descrito en la sección 7.3). Después de obtener una solución óptima para este modelo, se calcula

$$x_j = \sum_{k=1}^{n_j} x_{jk},$$

para $j = 1, 2, \ldots, n$, con el fin de identificar una solución óptima para el problema original de programación separable (o su aproximación lineal por partes).

Ejemplo. La Wyndor Glass Co. (vea la sección 3.1) recibió un pedido especial de artículos procesados a mano que se debe elaborar en las plantas 1 y 2 durante los próximos cuatro meses. Para cumplir con este pedido será necesario asignar algunos empleados de las brigadas de trabajo de los productos normales, por lo que el resto del personal tendrá que trabajar horas extra para utilizar toda la capacidad de producción de la maquinaria y equipo de la planta para elaborar estos productos. En particular, a fin de fabricar los dos nuevos productos normales que se analizaron en la sección 3.1, el tiempo extra tendrá que utilizar el último 25% de la capacidad disponible en la planta 1 para el producto 1 y el último 50% de la capacidad disponible en la planta 2 para el producto 2. El costo adicional del tiempo extra reducirá la ganancia de cada unidad del producto 1 de 3 a 2 dólares y del producto 2, de 5 dólares a 1 dólar, de lo que resultarán las *curvas de ganancia* de la figura 12.16, las cuales se ajustan a la forma del caso 1 de la figura 12.15.

La administración ha decidido usar tiempo extra en lugar de contratar más trabajadores durante esta situación temporal. Sin embargo, insiste en que se aprovechen por completo las brigadas de trabajo de cada producto en tiempo normal antes de usar cualquier tiempo extra. Aún más, piensa que temporalmente se deben cambiar las tasas de producción actuales ($x_1 = 2$ para el producto 1 y $x_2 = 6$ para el producto 2), si esta medida mejora el rendimiento total. Por todo ello, ha girado instrucciones al equipo de IO para revisar los productos 1 y 2 y determinar la nueva mezcla de productos más redituable durante los próximos cuatro meses.

[17] Si una o más de las $f_j(x_j)$ son ya funciones lineales, $f_j(x_j) = c_j x_j$, entonces $n_j = 1$, de manera que ninguna de estas sustituciones se hará para j.

[18] Se puede encontrar un algoritmo especializado muy eficiente para resolver este problema en R. Fourer: "A Specialized Algorithm for Piecewise-Linear Programming III: Computational Analysis and Applications", en *Mathematical Programming*, **53:** 213-235, 1992. Vea también A. M. Geoffrion: "Objective Function Approximations in Mathematical Programming", en *Mathematical Programming*, **13:** 23-37, 1977.

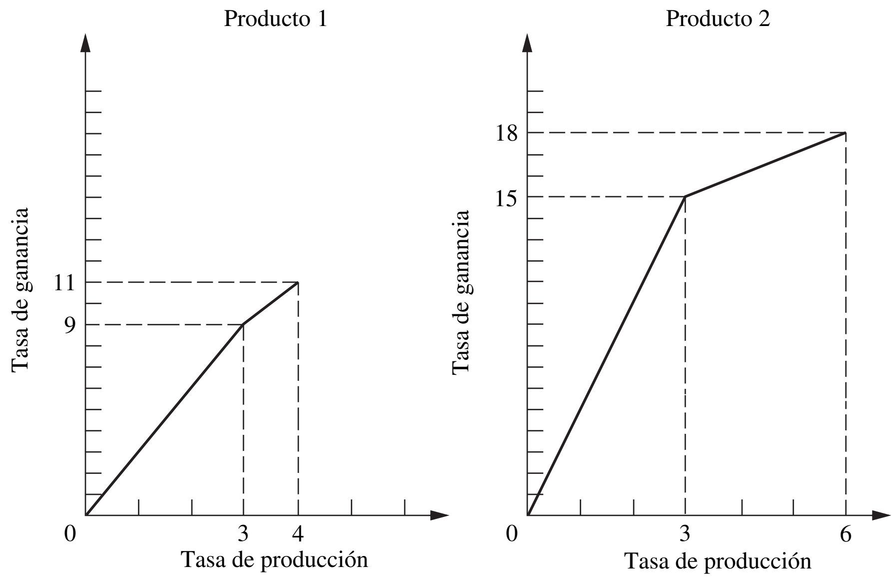


FIGURA 12.16
Datos de ganancia durante los próximos 4 meses de la Wyndor Glass Co.

Formulación. Para refrescar la memoria, el modelo de programación lineal del problema original de la Wyndor Glass Co., de la sección 3.1 es:

$$\text{Maximizar} \qquad Z = 3x_1 + 5x_2,$$

sujeta a

$$\begin{aligned} x_1 \qquad &\leq 4 \\ 2x_2 &\leq 12 \\ 3x_1 + 2x_2 &\leq 18 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

Ahora se debe modificar este modelo para que se ajuste a la nueva situación que se acaba de describir. Para esto, sea $x_1 = x_{1R} + x_{1O}$ la tasa de elaboración del producto 1, en donde x_{1R} es la tasa de producción alcanzada con tiempos normales de trabajo y x_{1O} es la tasa de producción incremental al usar tiempo extra. Defina $x_2 = x_{2R} + x_{2O}$ de la misma manera para el producto 2. Por tanto, en la notación general del modelo de programación lineal para programación separable que se presentó antes del ejemplo, $n = 2$, $n_1 = 2$ y $n_2 = 2$. Al sustituir los datos de la figura 12.16 —incluso las tasas máximas de producción en tiempo normal y tiempo extra— en este modelo general se obtiene el modelo específico para esta aplicación. En particular, el nuevo problema de programación lineal trata de determinar los valores de x_{1R}, x_{1O}, x_{2R} y x_{2O} para

$$\text{Maximizar} \qquad Z = 3x_{1R} + 2x_{1O} + 5x_{2R} + x_{2O},$$

sujeta a

$$\begin{aligned} x_{1R} + x_{1O} \qquad\qquad\qquad &\leq 4 \\ 2(x_{2R} + x_{2O}) &\leq 12 \\ 3(x_{1R} + x_{1O}) + 2(x_{2R} + x_{2O}) &\leq 18 \\ x_{1R} \leq 3, \qquad x_{1O} \leq 1, \qquad x_{2R} \leq 3, \qquad x_{2O} &\leq 3 \end{aligned}$$

y

$$x_{1R} \geq 0, \qquad x_{1O} \geq 0, \qquad x_{2R} \geq 0, \qquad x_{2O} \geq 0.$$

(Observe que las restricciones de cota superior del penúltimo renglón del modelo convierten a las primeras dos restricciones funcionales en *redundantes*, de manera que estas dos ecuaciones se pueden eliminar.)

Sin embargo, existe un factor importante que no se toma en cuenta en forma explícita en esta formulación. En particular, no existe nada en el modelo que requiera que se utilice todo el tiempo normal disponible, antes de emplear cualquier tiempo extra para elaborar ese producto. En otras palabras, puede ser factible tener $x_{1O} > 0$ aun cuando $x_{1R} < 3$ y tener $x_{2O} > 0$ aun cuando $x_{2R} < 3$. Tal solución sería inaceptable para la administración. (La prohibición de esta solución conduce a la *restricción especial* de la que se habló antes en esta misma sección.)

Como consecuencia, se llega a la *propiedad esencial de programación separable*. Aun cuando este modelo no toma en cuenta este factor de manera explícita, ¡lo hace en forma implícita! Aunque el modelo tenga un exceso de soluciones "factibles" que en realidad son inaceptables, *garantiza* que cualquier solución *óptima* será legítima, es decir, que no sustituirá tiempo de trabajo normal disponible por horas extra. (En este caso el razonamiento es análogo al del método de la *gran M* que se presentó en la sección 4.6, en el que también se permiten soluciones factibles adicionales, pero *no óptimas*, por conveniencia para trabajar.) En efecto, se puede aplicar con seguridad el método símplex a este modelo para encontrar la mezcla de productos más redituable. La razón es doble. Primero, las dos variables de decisión para cada producto *siempre* aparecen juntas como una *suma*, $x_{1R} + x_{1O}$ o $x_{2R} + x_{2O}$, en *cada* restricción funcional distinta de las restricciones de cota superior sobre las variables individuales. Por tanto, *siempre* es posible convertir una solución factible no aceptable en una aceptable que tenga las mismas tasas de producción totales, $x_1 = x_{1R} + x_{1O}$ y $x_2 = x_{2R} + x_{2O}$, con sólo reemplazar la producción en tiempo extra por la máxima producción normal. Segundo, la producción en tiempo extra es menos redituable que la normal —es decir, la pendiente de cada curva de ganancia de la figura 12.16 es una función monótona *decreciente* de la tasa de producción—, de modo que al convertir una solución factible no aceptable en una aceptable *debe* incrementarse la tasa total de la ganancia Z. En consecuencia, cualquier solución factible que usa tiempo extra de producción para un producto, cuando todavía se dispone de tiempo normal, *no puede* ser óptima respecto del modelo.

Por ejemplo, considere la solución factible inaceptable $x_{1R} = 1$, $x_{1O} = 1$, $x_{2R} = 1$, $x_{2O} = 3$, que proporciona una ganancia total $Z = 13$. La forma aceptable de lograr las mismas tasas de producción $x_1 = 2$ y $x_2 = 4$ es $x_{1R} = 2$, $x_{1O} = 0$, $x_{2R} = 3$, $x_{2O} = 1$. Esta última solución también es factible pero hace que el valor de Z aumente en $(3 - 2)(1) + (5 - 1)(2) = 9$ hasta una tasa total de ganancia $Z = 22$.

De manera similar, la solución óptima para este modelo resulta ser $x_{1R} = 3$, $x_{1O} = 1$, $x_{2R} = 3$, $x_{2O} = 0$, que es una solución factible aceptable.

En la sección Worked Examples de la página en internet de este libro se presenta **otro ejemplo** que ilustra la aplicación de la programación separable.

Extensiones

Hasta ahora, la presentación se ha centrado en el caso especial de programación separable, en el que la única función no lineal es la función objetivo $f(\mathbf{x})$. Ahora considere en forma breve el caso general en el que las funciones de restricción $g_i(\mathbf{x})$ no tienen que ser lineales, pero son convexas y separables, de manera que cada $g_i(\mathbf{x})$ se puede expresar como una suma de funciones de variables individuales

$$g_i(\mathbf{x}) = \sum_{j=1}^{n} g_{ij}(x_j),$$

donde cada $g_{ij}(x_j)$ es una función *convexa*. Una vez más, cada una de estas funciones se puede aproximar tanto como se quiera por una función *lineal por partes* (si no está ya en esa forma). La nueva restricción es que para cada variable x_j ($j = 1, 2, \dots, n$), todas las aproximaciones lineales por partes de las funciones de esa variable $[f_j(x_j), g_{1j}(x_j), \dots, g_{mj}(x_j)]$ deben tener los *mismos* límites de intervalos, para poder usar las mismas variables nuevas $(x_{j1}, x_{j2}, \dots, x_{jn_j})$ para todas las funciones lineales por partes. Esta formulación conduce a un modelo de programación lineal igual al que

se obtiene para el caso especial, excepto que para cada i y j, las variables x_{jk} tienen coeficientes distintos en la restricción i [donde estos coeficientes son las pendientes correspondientes de la función lineal por partes que aproxima a $g_{ij}(x_j)$]. Como $g_{ij}(x_j)$ tiene que ser convexa, en principio la misma lógica implica que debe cumplirse la propiedad esencial de programación separable. (Vea el problema 12.8-6*b*.)

Una desventaja de la aproximación de funciones por medio de funciones lineales por partes, descrita en esta sección, es que sólo se logra una buena aproximación con un número muy grande de segmentos de recta (variables), mientras que una malla fina de intervalos nada más se necesita en la vecindad inmediata de la solución óptima. Se han desarrollado[19] enfoques más elaborados, que usan una sucesión de funciones lineales por partes de *dos segmentos*, para obtener *aproximaciones cada vez más cercanas* en esta vecindad inmediata. Este tipo de enfoques tiende a ser más rápido y más exacto cuando se desea aproximar una solución óptima.

12.9 PROGRAMACIÓN CONVEXA

Se han estudiado algunos casos especiales de programación convexa en las secciones 12.4 y 12.5 (problemas no restringidos), 12.7 (función objetivo cuadrática con restricciones lineales) y 12.8 (funciones separables). También se vio, en la sección 12.6, una parte de la teoría del caso general (condiciones necesarias y suficientes para la optimalidad). En esta sección se presentarán de manera breve algunos tipos de enfoques que se utilizan para resolver el problema general de programación convexa [donde la función objetivo $f(\mathbf{x})$ que se debe maximizar es cóncava y las funciones de restricción $g_i(\mathbf{x})$ son convexas] y después se presentará un ejemplo de un algoritmo para programación convexa.

No existe un algoritmo estándar único que se pueda usar siempre para resolver problemas de programación convexa. Se han desarrollado muchos algoritmos diferentes, cada uno con ventajas y desventajas, y la investigación continúa activa en esta área. En términos generales, la mayor parte de estos algoritmos cae dentro de alguna de las tres categorías siguientes.

Una categoría la constituyen los **algoritmos de gradiente**, en los que se modifica de alguna manera el procedimiento de búsqueda del gradiente de la sección 12.5 para evitar que la trayectoria de búsqueda penetre la frontera de restricción. Por ejemplo, un método popular es el *método de gradiente reducido generalizado* (GRG). El Excel Solver emplea el método GRG para resolver problemas de programación convexa. (Como se explica en la sección siguiente, el Premium Solver también incluye una opción evolutiva que fue diseñada para tratar con problemas de programación *no convexa*.)

La segunda categoría, los **algoritmos secuenciales no restringidos**, incluye los métodos de *función de finalización* y de *función barrera*. Estos algoritmos convierten el problema de optimización restringida original en una sucesión de problemas de *optimización no restringida* cuyas soluciones óptimas convergen a la solución óptima del problema original. Cada uno de estos problemas de optimización no restringida se puede resolver por medio del procedimiento de búsqueda del gradiente de la sección 12.5. Esta conversión se logra al incorporar las restricciones a una función de penalización (o función barrera) que se resta de la función objetivo, con el fin de imponer un castigo grande a la violación de cualquier restricción —o aun al hecho de estar cerca de los límites—. En la última parte de esta sección se describirá un algoritmo de la década de 1960, llamado **técnica de minimización no restringida secuencial** (o **SUMT**, por sus siglas en inglés), que fue pionero en esta categoría de algoritmos. (SUMT también ayudó a motivar algunos de los *métodos de punto interior* para programación lineal.)

La tercera categoría, los **algoritmos de aproximación secuencial**, incluye métodos de *aproximación lineal* y *aproximación cuadrática*. Estos algoritmos sustituyen la función objetivo no lineal por una sucesión de aproximaciones lineales o cuadráticas. Para problemas de optimización linealmente restringidos, estas aproximaciones permiten la aplicación repetida de los algoritmos de programación lineal o cuadrática. Este trabajo se logra mediante otro análisis que conduce a una sucesión de soluciones que convergen a una solución óptima para el problema original. Aunque estos algoritmos son adecuados en especial para problemas de optimización linealmente restringi-

[19] R. R. Meyer: "Two-Segment Separable Programming", en *Management Science*, **25:** 385-395, 1979.

da, algunos se pueden extender a problemas con funciones de restricción no lineales, si se usan las aproximaciones lineales adecuadas.

Como un ejemplo de algoritmo de *aproximación secuencial* se presentará el **algoritmo de Frank-Wolfe**[20] para el caso de programación convexa *linealmente restringida* (de manera que las restricciones son $\mathbf{Ax} \leq \mathbf{b}$ y $\mathbf{x} \geq \mathbf{0}$, en forma matricial). Este procedimiento es bastante directo: combina aproximaciones lineales de la función objetivo (que permiten usar el método símplex) con el procedimiento para la optimización no restringida de una variable (como se describió en la sección 12.4).

Algoritmo de aproximación lineal secuencial (Frank-Wolfe)

Dada una solución de prueba factible $\mathbf{x}'$, la aproximación lineal que se usa para la función objetivo $f(\mathbf{x})$ es la expansión de primer orden por series de Taylor, de $f(\mathbf{x})$ alrededor de $\mathbf{x} = \mathbf{x}'$, es decir,

$$f(\mathbf{x}') \approx f(\mathbf{x}') + \sum_{j=1}^{n} \frac{\partial f(\mathbf{x}')}{\partial x_j}(x_j - x_j') = f(\mathbf{x}') + \nabla f(\mathbf{x}')(\mathbf{x} - \mathbf{x}'),$$

donde estas derivadas parciales se evalúan en $\mathbf{x} = \mathbf{x}'$. Como $f(\mathbf{x}')$ y $\nabla f(\mathbf{x}')\mathbf{x}'$ tienen valores fijos, se pueden eliminar para obtener una función objetivo lineal equivalente,

$$g(\mathbf{x}) = \nabla f(\mathbf{x}')\mathbf{x} = \sum_{j=1}^{n} c_j x_j, \qquad \text{donde } c_j = \frac{\partial f(\mathbf{x})}{\partial x_j} \qquad \text{en } \mathbf{x} = \mathbf{x}'.$$

Después se aplica el método símplex (o el procedimiento gráfico, si $n = 2$) al problema de programación lineal (LP) que resulta [maximizar $g(\mathbf{x})$ sujeta a las restricciones originales $\mathbf{Ax} \leq \mathbf{b}$ y $\mathbf{x} \geq \mathbf{0}$] a fin de obtener *su* solución óptima x_{LP}. Observe que la función objetivo lineal necesariamente se incrementa cuando se hace un recorrido sobre el segmento de recta desde $\mathbf{x}'$ hasta $\mathbf{x}_{LP}$ (que se encuentra en la frontera de la región factible); sin embargo, la aproximación lineal puede no ser buena para una $\mathbf{x}$ lejana a $\mathbf{x}'$, por lo que quizá la función objetivo *no lineal* puede no continuar creciendo durante toda la trayectoria desde $\mathbf{x}'$ hasta $\mathbf{x}_{LP}$. Así, en lugar de aceptar $\mathbf{x}_{LP}$ como la siguiente solución de prueba, se elige el punto sobre este segmento de recta que maximiza la función objetivo no lineal. Este punto se puede encontrar si se realiza una *búsqueda en una dimensión* mediante el procedimiento de la sección 12.4, donde la única variable para propósitos de esta búsqueda es la fracción t de la distancia total de $\mathbf{x}'$ a $\mathbf{x}_{LP}$. En consecuencia, este punto se convierte en una nueva solución de prueba para iniciar la siguiente iteración del algoritmo, como se acaba de describir. La sucesión de soluciones de prueba que generan las iteraciones converge hacia una solución óptima para el problema original, de manera que el algoritmo se detiene en cuanto las soluciones de prueba están lo suficientemente cerca entre sí como para poder concluir que se llegó a esta solución óptima.

Resumen de algoritmo de Frank-Wolfe

Paso inicial: se encuentra una solución de prueba factible inicial $\mathbf{x}^{(0)}$, por ejemplo, al aplicar los procedimientos de programación lineal para encontrar una solución BF. Se hace $k = 1$.

Iteración k:

1. Para $j = 1, 2, \ldots, n$, se evalúa

$$\frac{\partial f(\mathbf{x})}{\partial x_j} \qquad \text{en } \mathbf{x} = \mathbf{x}^{(k-1)}$$

y se iguala c_j a este valor.

[20] M. Frank y P. Wolfe, “An Algorithm for Quadratic Programming”, en *Naval Research Logistics Quarterly*, **3:** 95-110, 1956. Aunque originalmente se diseñó para programación cuadrática, este algoritmo se adapta con facilidad al caso de una función objetivo cóncava general que se estudia aquí.

2. Se encuentra una solución óptima $\mathbf{x}_{LP}^{(k)}$ para el siguiente problema de programación lineal.

$$\text{Maximizar} \qquad g(\mathbf{x}) = \sum_{j=1}^{n} c_j x_j,$$

sujeta a

$$\mathbf{Ax} \leq \mathbf{b} \qquad \text{y} \qquad \mathbf{x} \geq \mathbf{0}.$$

3. Para la variable t ($0 \leq t \leq 1$), se establece

$$h(t) = f(\mathbf{x}) \qquad \text{para } \mathbf{x} = \mathbf{x}^{(k-1)} + t(\mathbf{x}_{LP}^{(k)} - \mathbf{x}^{(k-1)}),$$

de manera que $h(t)$ proporciona el valor de $f(\mathbf{x})$ sobre el segmento de recta entre $\mathbf{x}^{(k-1)}$ (donde $t = 0$) y $\mathbf{x}_{LP}^{(k)}$ (donde $t = 1$). Se aplica algún procedimiento como el de búsqueda en una dimensión (vea la sección 12.4) para maximizar $h(t)$ en $0 \leq t \leq 1$ y se establece $\mathbf{x}^{(k)}$ igual a la $\mathbf{x}$ correspondiente. Se pasa a la regla de detención.

Regla de detención: si $\mathbf{x}^{(k-1)}$ y $\mathbf{x}^{(k)}$ están lo suficientemente cerca, el proceso se detiene y se usa $\mathbf{x}^{(k)}$ (o una extrapolación de $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k-1)}, \mathbf{x}^{(k)}$) como la estimación de una solución óptima. De otra manera se hace $k = k + 1$ y se realiza otra iteración.

Ahora se ilustrará este procedimiento.

Ejemplo. Considere el siguiente problema de programación convexa linealmente restringido;

$$\text{Maximizar} \qquad f(\mathbf{x}) = 5x_1 - x_1^2 + 8x_2 - 2x_2^2,$$

sujeta a

$$3x_1 + 2x_2 \leq 6$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

Observe que

$$\frac{\partial f}{\partial x_1} = 5 - 2x_1, \qquad \frac{\partial f}{\partial x_2} = 8 - 4x_2,$$

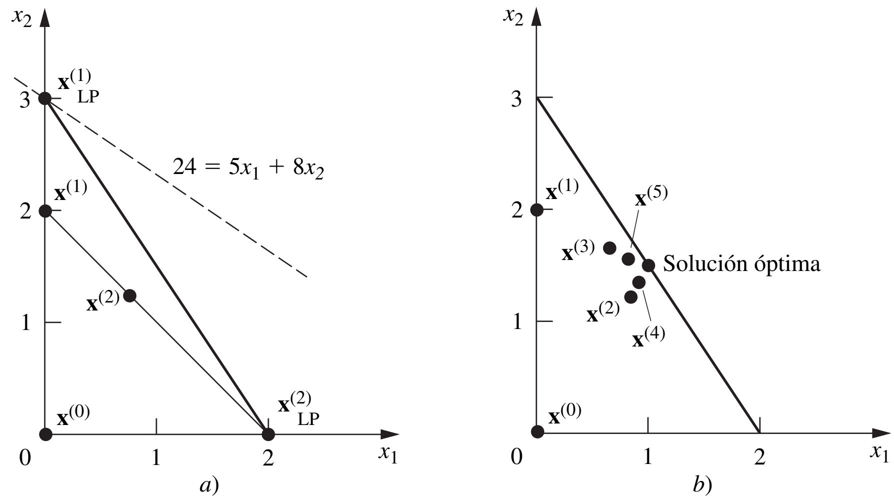
por lo que el máximo *no restringido* $\mathbf{x} = (\frac{5}{2}, 2)$, viola las restricciones funcionales. Por tanto, se necesita trabajar un poco más para encontrar el máximo *restringido*.

Iteración 1: como $\mathbf{x} = (0, 0)$, es claro que es factible (y corresponde a la solución inicial BF de las restricciones de programación lineal), se elige como la solución de prueba inicial $\mathbf{x}^{(0)}$ para el algoritmo de Frank-Wolfe. Al sustituir $x_1 = 0$ y $x_2 = 0$ en las expresiones de las derivadas parciales se obtiene $c_1 = 5$ y $c_2 = 8$, con lo que $g(\mathbf{x}) = 5x_1 + 8x_2$ es la primera aproximación lineal de la función objetivo. Con la solución gráfica de este problema de programación lineal (vea la figura 12.17*a*) se llega a $\mathbf{x}_{LP}^{(1)} = (0, 3)$. En la parte 3 del paso iterativo, los puntos en el segmento de recta que une a (0, 0) y (0, 3), que se muestran en la figura 12.17*a*, se expresan por

$$\begin{aligned}(x_1, x_2) &= (0, 0) + t[(0, 3) - (0, 0)] \qquad \text{para } 0 \leq t \leq 1 \\ &= (0, 3t)\end{aligned}$$

como se muestra en la sexta columna de la tabla 12.6. Esta expresión da entonces

$$\begin{aligned}h(t) &= f(0, 3t) = 8(3t) - 2(3t)^2 \\ &= 24t - 18t^2,\end{aligned}$$



■ **FIGURA 12.17**
Ilustración del algoritmo de Frank-Wolfe.

de modo que el valor $t = t^*$ que maximiza $h(t)$ en el intervalo $0 \leq t \leq 1$ se puede obtener, en este caso, al establecer

$$\frac{dh(t)}{dt} = 24 - 36t = 0,$$

por lo que $t^* = \frac{2}{3}$. Este resultado conduce a la siguiente solución de prueba

$$\mathbf{x}^{(1)} = (0, 0) + \frac{2}{3}[(0, 3) - (0, 0)]$$
$$= (0, 2),$$

lo cual completa la primera iteración.

Iteración 2: para esbozar los cálculos que conducen a los resultados del segundo renglón de la tabla 12.6, observe que $\mathbf{x}^{(1)} = (0, 2)$ da

$$c_1 = 5 - 2(0) = 5,$$
$$c_2 = 8 - 4(2) = 0.$$

Al resolver el problema en una gráfica con la función objetivo $g(\mathbf{x}) = 5x_1$ en la región factible de la figura 12.17*a*, se obtiene $\mathbf{x}^{(2)}_{LP} = (2, 0)$. De esta forma, la expresión del segmento de recta entre $\mathbf{x}^{(1)}$ y $\mathbf{x}^{(2)}_{LP}$ (vea la figura 12.17*a*) es

$$\mathbf{x} = (0, 2) + t[(2, 0) - (0, 2)]$$
$$= (2t, 2 - 2t),$$

de manera que

$$h(t) = f(2t, 2 - 2t)$$
$$= 5(2t) - (2t)^2 + 8(2 - 2t) - 2(2 - 2t)^2$$
$$= 8 + 10t - 12t^2.$$

■ **TABLA 12.6 Aplicación del algoritmo de Frank-Wolfe al ejemplo**

k	$\mathbf{x}^{(k-1)}$	c_1	c_2	$\mathbf{x}^{(k)}_{LP}$	$\mathbf{x}$ para $h(t)$	$h(t)$	t^*	$\mathbf{x}^{(k)}$
1	(0, 0)	5	8	(0, 3)	$(0, 3t)$	$24t - 18t^2$	$\frac{2}{3}$	(0, 2)
2	(0, 2)	5	0	(2, 0)	$(2t, 2 - 2t)$	$8 + 10t - 12t^2$	$\frac{5}{12}$	$\left(\frac{5}{6}, \frac{7}{6}\right)$

Si se establece

$$\frac{dh(t)}{dt} = 10 - 24t = 0$$

se llega a $t^* = \frac{5}{12}$. Entonces

$$\mathbf{x}^{(2)} = (0, 2) + \frac{5}{12}[(2, 0) - (0, 2)]$$

$$= \left(\frac{5}{6}, \frac{7}{6}\right),$$

con lo que se completa la segunda iteración.

En la figura 12.17*b* se puede observar cómo las soluciones de prueba adquieren valores de puntos alternados entre dos trayectorias que parecen cruzarse aproximadamente en el punto $\mathbf{x} = (1, \frac{3}{2})$. En realidad, este punto es la solución óptima, como se puede verificar si se aplican las condiciones KKT de la sección 12.6.

Este ejemplo ilustra una característica común del algoritmo de Frank-Wolfe: que las soluciones de prueba alternan entre dos (o más) trayectorias. Cuando esto ocurre, se pueden extrapolar las trayectorias al punto aproximado en que se cruzan, para estimar una solución óptima. Esta estimación tiende a ser mejor que usar la última solución de prueba generada. Esto se debe a que la convergencia de las soluciones de prueba hacia la solución óptima tiende a ser lenta y por ello es posible que la última solución esté todavía alejada del óptimo.

Si desea ver **otro ejemplo** de aplicación del algoritmo de Frank-Wolfe, en la sección Worked Examples del sitio en internet de este libro se incluye uno. En el OR Tutor también se proporciona un **ejemplo adicional**. Asimismo en el IOR Tutorial se incluye un algoritmo interactivo para este algoritmo.

Algunos otros algoritmos

Se debe hacer hincapié en que el algoritmo de Frank-Wolfe es sólo un ejemplo de los algoritmos de aproximación secuencial. Muchos de ellos utilizan aproximaciones *cuadráticas* en lugar de *lineales* en cada iteración, porque la aproximación cuadrática proporciona un ajuste mucho mejor al problema original y por lo mismo permite que la sucesión de soluciones converja mucho más rápido hacia la solución óptima que en el caso de la figura 12.17*b*. Por esta razón, aun cuando el uso de los métodos de aproximación lineal secuencial, como el algoritmo de Frank-Wolfe, sea bastante sencillo y directo, en general, para las aplicaciones reales se prefieren los *métodos de aproximación cuadrática*. Entre éstos, los que más se usan son los métodos *cuasi-Newton* (o de *variables métricas*), que calculan una aproximación cuadrática a la curvatura de una función no lineal sin calcular en forma explícita las segundas derivadas (parciales). (En los problemas de optimización linealmente restringidos, esta función no lineal es justo la función objetivo, mientras que si se tienen restricciones no lineales, se trata de la función de Lagrange descrita en el apéndice 3.) Algunos algoritmos cuasinewtonianos no formulan ni dan una solución aproximada del problema de programación cuadrática en cada iteración, sino que incorporan algunos de los ingredientes básicos de los *algoritmos de gradiente*. (Vea la referencia seleccionada 2 para consultar detalles adicionales acerca de los algoritmos de aproximación secuencial.)

Ahora se pasará de los algoritmos de aproximación secuencial a los *algoritmos no restringidos secuenciales*. Como se mencionó al inicio de la sección, los algoritmos del último tipo resuelven el problema de optimización restringida original en lugar de resolver una secuencia de problemas de optimización *no restringida*.

Un procedimiento de este tipo que se utiliza de manera profusa desde su origen en la década de 1960 es la *técnica secuencial de minimización no restringida* (o SUMT, *sequential unconstrained minimization technique*).[21] En realidad, existen dos versiones de SUMT, una de las cuales es un

[21] Vea la referencia seleccionada 1.

algoritmo *de punto exterior* que utiliza soluciones *no factibles* al mismo tiempo que una *función de penalización* para forzar la convergencia hacia la región factible. Se describirá la otra versión, que es un algoritmo *de punto interior* que trabaja en forma directa con soluciones *factibles* y, de manera simultánea, utiliza una *función barrera* para forzarlas a mantenerse dentro de esa región factible. Aunque la técnica SUMT en un principio se presentó como una técnica de minimización, se convertirá aquí en una técnica de maximización, a fin de que sea congruente con el resto del capítulo. En consecuencia, se continuará con el supuesto de que el problema se encuentra en la forma dada al principio y que todas las funciones son diferenciables.

Técnica secuencial de minimización no restringida (SUMT)

Como su nombre lo indica, SUMT sustituye el problema original por una *sucesión* de problemas de optimización *no restringida*, cuyas soluciones *convergen* hacia una solución (máximo local) del problema original. Este enfoque es muy atractivo, pues es mucho más sencillo resolver los problemas de optimización no restringida (vea la sección 12.5) que los que tienen restricciones. Cada uno de los problemas no restringidos de esta sucesión incluye la selección de un valor estrictamente positivo (cada vez más pequeño) de un escalar r y después la obtención de $\mathbf{x}$ para

$$\text{Maximizar} \quad P(\mathbf{x}; r) = f(\mathbf{x}) - rB(\mathbf{x}).$$

Aquí $B(\mathbf{x})$ es una **función barrera** que tiene las propiedades (para las $\mathbf{x}$ factibles del problema original) siguientes:

1. $B(\mathbf{x})$ es *pequeño* cuando $\mathbf{x}$ está *lejos* de la frontera de la región factible.
2. $B(\mathbf{x})$ es *grande* cuando $\mathbf{x}$ está *cerca* de la frontera de la región factible.
3. $B(\mathbf{x}) \to \infty$ cuando la distancia desde la frontera (más cercana) de la región factible $\to 0$.

Así, si se inicia el procedimiento de búsqueda con una solución de prueba inicial *factible* y después se intenta incrementar $P(\mathbf{x}; r)$, entonces $B(\mathbf{x})$ proporciona una *barrera* que evita que la búsqueda cruce (o llegue a) la frontera de la región factible del problema original.

La elección más natural para $B(\mathbf{x})$ es

$$B(\mathbf{x}) = \sum_{i=1}^{m} \frac{1}{b_i - g_i(\mathbf{x})} + \sum_{j=1}^{n} \frac{1}{x_j}.$$

Observe que para valores factibles de $\mathbf{x}$, el denominador de cada término es proporcional a la distancia de $\mathbf{x}$ a la frontera de restricciones de la restricción funcional o de no negatividad correspondiente. En consecuencia, *cada* término es un *término de rechazo de frontera* que tiene las tres propiedades anteriores con respecto a esta frontera de restricción específica. Otra característica atractiva de esta $B(\mathbf{x})$ es que cuando se satisfacen todos los supuestos *de programación convexa*, $P(\mathbf{x}; r)$ es una función *cóncava*.

Como $B(\mathbf{x})$ mantiene la búsqueda lejos de la frontera de la región factible, una pregunta natural es: ¿qué pasa si la solución que se quiere está ahí? Esta preocupación es la razón por la que SUMT incluye la solución de una *sucesión* de estos problemas de optimización no restringida para valores cada vez más pequeños de r que se acercan a cero (donde la solución de prueba final de cada uno se convierte en la solución de prueba inicial del siguiente). Por ejemplo, cada nueva r se puede obtener a partir de la anterior multiplicando por una constante $\theta(0 < \theta < 1)$, en donde un valor usual es $\theta = 0.01$. A medida que r se acerca a 0, $P(\mathbf{x}; r)$ se acerca a $f(\mathbf{x})$, de manera que el máximo local correspondiente de $P(\mathbf{x}; r)$ converge hacia un máximo local del problema original. Por ello, sólo es necesario resolver problemas de optimización no restringida, hasta que se puedan extrapolar sus soluciones a esta solución límite.

¿Cuántos problemas son suficientes para permitir esta extrapolación? Cuando el problema original satisface los supuestos de programación convexa, se cuenta con información útil como guía en esta decisión. En particular, si $\bar{x}$ es un máximo global de $P(\mathbf{x}; r)$, entonces

$$f(\bar{\mathbf{x}}) \leq f(\mathbf{x}^*) \leq f(\bar{\mathbf{x}}) + rB(\bar{\mathbf{x}}),$$

donde $\mathbf{x}^*$ es la solución *óptima* (desconocida) para el problema original. Entonces, $rB(\bar{\mathbf{x}})$ es el *error máximo* (en el valor de la función objetivo) que puede resultar al usar $\bar{\mathbf{x}}$ para aproximar $\mathbf{x}^*$,

y este error disminuye al extrapolar a un valor mayor a $\bar{\mathbf{x}}$ para incrementar $f(\mathbf{x})$. Si de antemano se establece la *tolerancia del error*, entonces puede detenerse el proceso tan pronto como $rB(\bar{\mathbf{x}})$ sea menor que esta cantidad.

Resumen de la SUMT

Paso inicial: se identifica una solución de prueba inicial *factible* $\mathbf{x}^{(0)}$ que no esté en la frontera de la región factible. Se hace $k = 1$ y se eligen valores adecuados estrictamente positivos para la r inicial y para $\theta < 1$ (por ejemplo, $r = 1$ y $\theta = 0.01$).[22]

Iteración k: si se comienza con $\mathbf{x}^{(k-1)}$, se aplica el procedimiento de búsqueda del gradiente descrito en la sección 12.5 (o un método similar) para encontrar un máximo local, $\mathbf{x}^{(k)}$ de

$$P(\mathbf{x};\, r) = f(\mathbf{x}) - r\left[\sum_{i=1}^{m} \frac{1}{b_i - g_i(\mathbf{x})} + \sum_{j=1}^{n} \frac{1}{x_j}\right].$$

Regla de detención: si el cambio de $\mathbf{x}^{(k-1)}$ a $\mathbf{x}^{(k)}$ es despreciable, el proceso se detiene y se usa $\mathbf{x}^{(k)}$ (o una extrapolación de $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k-1)}, \mathbf{x}^{(k)}$) como la estimación de un máximo local del problema original. De otra manera, se establece $k = k + 1$ y $r = \theta r$ y se realiza otra iteración.

Por último, observe que la técnica SUMT se puede extender para manejar restricciones de *igualdad*, $g_i(\mathbf{x}) = b_i$. Una forma estándar es la siguiente. Para cada restricción de igualdad,

$$\frac{-[b_i - g_i(\mathbf{x})]^2}{\sqrt{r}} \qquad \text{sustituye a} \qquad \frac{-r}{b_i - g_i(\mathbf{x})}$$

en la expresión de $P(\mathbf{x};\, r)$ que se dio en el resumen de la técnica; después se usa el mismo procedimiento. El numerador, $-[b_i - g_i(\mathbf{x})]^2$, impone una gran penalización para la desviación grande de la satisfacción de la restricción de igualdad, y después el denominador aumenta en forma tremenda esta penalización a medida que r disminuye a una cantidad muy pequeña, lo que hace que la secuencia de soluciones de prueba converjan hacia el punto que satisface la restricción.

La SUMT se utiliza de manera profusa por su simplicidad y versatilidad. Sin embargo, quienes se dedican al análisis numérico han encontrado que es bastante propensa a la *inestabilidad numérica*, por lo que se aconseja tomar precauciones extremas. Para mayor información sobre este tema, igual que sobre análisis similares para otros algoritmos, vea la referencia 3.

Ejemplo Para ilustrar la SUMT considere el siguiente problema de dos variables.

$$\text{Maximizar} \qquad f(\mathbf{x}) = x_1 x_2,$$

sujeta a

$$x_1^2 + x_2 \leq 3$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

TABLA 12.7 Ilustración de la SUMT

k	r	$x_1^{(k)}$	$x_2^{(k)}$
0		1	1
1	1	0.90	1.36
2	10^{-2}	0.987	1.925
3	10^{-4}	0.998	1.993
		↓	↓
		1	2

[22] Un criterio razonable para elegir la r inicial es la que hace que $rB(\mathbf{x})$ sea de más o menos del mismo orden de magnitud que $f(\mathbf{x})$ para soluciones $\mathbf{x}$ que no se encuentren muy cerca de la frontera.

Aun cuando $g_1(\mathbf{x}) = x_1^2 + x_2$ es convexa (puesto que todos los términos son convexos), se trata de un problema de programación *no convexa* porque $f(\mathbf{x}) = x_1x_2$ *no* es cóncava (vea el apéndice 2). Sin embargo, el problema está tan cerca de ser un problema de programación convexa que necesariamente SUMT convergirá hacia una solución óptima en este caso. (En la siguiente sección se explicará más a fondo la programación no convexa, exposición que incluye el papel de la SUMT al enfrentar este tipo de problemas.)

En el paso inicial, $(x_1, x_2) = (1, 1)$ es una solución obvia que no está en la frontera de la región factible, de manera que se puede establecer $\mathbf{x}^{(0)} = (1, 1)$. Una elección razonable para r y θ es $r = 1$ y $\theta = 0.01$.

Para cada iteración,

$$P(\mathbf{x};\, r) = x_1x_2 - r\left(\frac{1}{3 - x_1^2 - x_2} + \frac{1}{x_1} + \frac{1}{x_2}\right).$$

Con $r = 1$, se aplica el procedimiento de búsqueda del gradiente a partir de (1, 1) para maximizar esta expresión, hasta llegar a $\mathbf{x}^{(1)} = (0.90, 1.36)$. Se establece $r = 0.01$ y se comienza de nuevo el procedimiento de búsqueda del gradiente a partir de (0.90, 1.36), y se llega entonces a $\mathbf{x}^{(2)} = (0.983, 1.933)$. Una iteración más con $r = 0.01(0.01) = 0.0001$ conduce de $\mathbf{x}^{(2)}$ a $\mathbf{x}^{(3)} = (0.998, 1.994)$. Es claro que esta sucesión de puntos, que se resume en la tabla 12.7, es convergente a (1, 2). Si se aplican las condiciones KKT a esta solución, se verifica que sin duda satisface la condición necesaria para la optimalidad. Un análisis gráfico demuestra que $(x_1, x_2) = (1, 2)$ es, en realidad, un máximo global (vea el problema 12.9-13*b*).

En este problema no existen máximos locales distintos de $(x_1, x_2) = (1, 2)$, por lo que si se vuelve a aplicar la técnica SUMT a partir de otras soluciones de prueba factibles iniciales, siempre se llegará a esta misma solución.[23]

En la sección Worked Examples del sitio en internet de este libro se proporciona **otro ejemplo** que ilustra la aplicación de la SUMT a un problema de programación convexa en forma de minimización. También en el OR Tutor se puede ver un **ejemplo adicional**. Asimismo, se incluye un procedimiento automático para ejecutar la SUMT.

Algunas opciones de software para programación convexa

Como se indicó en la sección 12.7, tanto Excel como LINGO pueden resolver problemas de programación convexa, no así LINDO y CPLEX, excepto en el caso especial de programación cuadrática (cuyo ejemplo se incluye en esta sección). Los detalles de este ejemplo se dan en los archivos de Excel y LINGO/LINDO para este capítulo en el OR Courseware. La versión profesional de MPL maneja un gran número de solucionadores que incluyen algunos que pueden resolver problemas de programación convexa. Uno de ellos, llamado CONOPT, se incorpora a la versión de MPL que contiene el sitio en internet de este libro. Los ejemplos de programación convexa que se presentan en el archivo de este capítulo de MPL se resolvieron con este solucionador después de establecer el tipo de modelo no lineal (como se describió al final de la sección 12.7).

12.10 PROGRAMACIÓN NO CONVEXA

Los supuestos de programación convexa (la función $f(\mathbf{x})$ que debe maximizarse es *cóncava* y todas las restricciones $g_i(\mathbf{x})$ son *convexas*) son muy convenientes puesto que aseguran que cualquier *máximo local* es también un *máximo global*. (Si, por el contrario, el objetivo es minimizar $f(\mathbf{x})$, entonces la programación convexa supone que $f(\mathbf{x})$ es convexa, lo que asegura que un mínimo local también es un mínimo global.) Desafortunadamente, los problemas de programación no lineal que surgen en la práctica muchas veces no cumplen estos supuestos. ¿Qué tipo de enfoque conviene usar para manejar estos problemas de *programación no convexa*?

[23] La razón técnica es que $f(\mathbf{x})$ es una función (estrictamente) *cuasicóncava* que comparte la propiedad de las funciones cóncavas de que un máximo local es siempre un máximo global. Más información al respecto se encuentra en M. Avriel, W. E. Diewert, S. Schaible e I. Zang, *Generalized Concavity*, Plenum, Nueva York, 1985.

El reto de resolver problemas de programación no convexa

No existe una respuesta única para la pregunta anterior porque hay muchos tipos diferentes de problemas de programación no convexa. Algunos son mucho más difíciles de resolver que otros. Por ejemplo, un problema de maximización donde la función objetivo está cerca de ser *convexa* por lo general es mucho más difícil de resolver que uno donde la función objetivo es casi cóncava. (El ejemplo de la SUMT en la sección 12.9 ilustró un caso donde la función objetivo estaba tan cerca de ser cóncava que el problema podía ser tratado como si fuera un problema de programación convexa.) De manera similar, al tener una región factible que *no* es un conjunto convexo (porque algunas de las funciones $g_i(\mathbf{x})$ no son convexas) por lo general representa una complicación importante. Al tratar con funciones que no son diferenciables, o quizá ni siquiera continuas, también se presenta una complicación relevante.

La meta de una gran parte de la investigación que se realiza en la actualidad es desarrollar procedimientos eficientes de *optimización global* para encontrar una *solución óptima global* para diferentes tipos de problemas de programación no convexa, y se han hecho algunos progresos. Como ejemplo, LINDO Systems (cuyos procedimientos son LINDO, LINGO y What's Best!) ha incorporado un optimizador global en el solucionador avanzado que comparten algunos de sus productos de software. En particular, LINGO y What's Best! tienen una función de inicio múltiple para generar de manera automática un número de puntos de inicio para su solucionador de programación no lineal con el fin de encontrar con rapidez una buena situación. Si se selecciona la opción global, en seguida estos paquetes emplean el optimizador global. Este optimizador convierte un problema de programación no convexa (incluso aquellos cuya formulación involucra funciones lógicas como IF, AND, OR y NOT) en algunos subproblemas que son solturas de programación convexa de partes del problema original. Después se utiliza la técnica de ramificación y acotamiento para buscar de manera exhaustiva entre los subproblemas. Una vez que el procedimiento llega a su terminación, se garantiza que la función que se encuentre sea una solución óptima global. (La otra conclusión posible es que el problema no tenga soluciones factibles.) La versión para estudiantes de este optimizador global se incluye en la versión de LINGO que se proporciona en el sitio en internet de este libro. Sin embargo, está limitado a problemas relativamente pequeños (con un máximo de cinco variables). La versión profesional del optimizador global ha resuelto algunos problemas mucho más grandes.

De manera similar, en la actualidad MPL incluye un optimizador global llamado LGO. La versión para estudiantes de LGO está disponible como uno de los solucionadores de MPL que se incluye en el sitio en internet de este libro. LGO también se puede usar para resolver problemas de programación convexa.

Se ha probado con diversos enfoques para abordar la optimización global (como el incorporado a LINGO que se describe a continuación). No se intentará afrontar este tema avanzado de ninguna forma (vea la referencia 5 para mayores detalles). En lugar de ello se iniciará con un caso simple y después se introducirá un enfoque más general al final de la sección. La metodología se ilustrará mediante el uso de hojas de cálculo de Excel, pero también pueden emplearse otros paquetes de software.

Uso de Excel Solver para encontrar óptimos locales

A continuación se estudiarán los enfoques directos sobre tipos relativamente simples de problemas de programación no convexa. En particular se considerarán problemas (de maximización) donde la función objetivo es casi cóncava ya sea sobre toda la región factible o dentro de porciones significativas de ella. También se ignorará la complejidad adicional que implica tener funciones de restricción $g_i(\mathbf{x})$ no convexas al usar únicamente restricciones lineales. Se comenzará por ilustrar lo que se puede lograr sólo con aplicar algún algoritmo para programación no convexa a dichos problemas. Aunque se podría seleccionar alguno de esos algoritmos (como los descritos en la sección 12.9), se usará el algoritmo de programación convexa que emplea el Solver de Excel para problemas de programación no lineal.

Por ejemplo, considere el siguiente problema de programación no convexa de una variable.

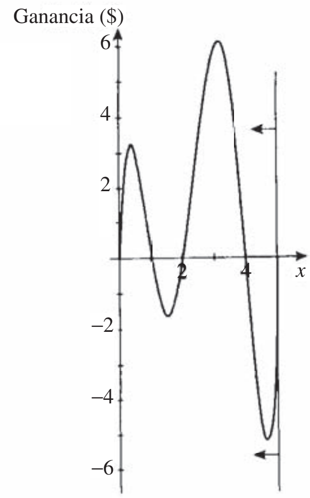
$$\text{Maximizar} \qquad Z = 0.5x^5 - 6x^4 + 24.5x^3 - 39x^2 + 20x,$$

sujeta a

$$x \leq 5$$
$$x \geq 0,$$

donde Z representa la ganancia en dólares. En la figura 12.18 se muestra una gráfica de la ganancia a través de la región factible que señala el elevado grado de no convexidad de esta función. Sin embargo, si esta gráfica no estuviera disponible, podría no ser claro de inmediato que éste no es un problema de programación convexa puesto que se requiere un pequeño análisis para verificar que la función objetivo no es cóncava en la región factible. Por lo tanto, suponga que Excel Solver, que está diseñado para resolver problemas de programación convexa, se aplica a este ejemplo. En la figura 12.19 se demuestra lo difícil que resulta para Excel Solver tratar de resolver este problema. El modelo se puede formular de manera directa en una hoja de cálculo, con x (C5) como la celda cambiante y la Ganancia (C8) como la celda objetivo. [Observe que la opción de Solver, *Assume Linear Model* (Asumir modelo lineal), no está seleccionada en este caso porque éste no es un modelo de programación lineal.] Cuando se introduce $x = 0$ como el valor inicial en la celda cambiante, la hoja de cálculo de la izquierda de la figura 12.19 muestra que el solver indica que $x = 0.371$ es la solución óptima con Ganancia = \$3.19. Sin embargo, si se introduce $x = 3$ como el valor inicial, como se muestra en la hoja de cálculo de en medio de dicha figura, Solver obtiene $x = 3.126$ como la solución óptima con Ganancia = \$6.13. Si se prueba otro valor inicial de $x = 4.7$ en la hoja de cálculo de la derecha, Solver ahora indica una solución óptima de $x = 5$ con Ganancia = \$0. ¿Qué sucede aquí?

En la figura 12.18 se puede encontrar una explicación de las dificultades de Solver con este problema. A partir de $x = 0$, la gráfica de ganancia en realidad asciende hasta un pico en $x = 0.371$, como se reporta en la hoja de cálculo de la izquierda de la figura 12.19. Si en lugar de eso se inicia en $x = 3$, la gráfica sube hasta un pico en $x = 3.126$ que es la solución que se encontró en la hoja de cálculo de en medio. Si se usa la hoja de cálculo de la derecha a partir de la solución de $x = 4.7$, la gráfica asciende hasta que alcanza la frontera impuesta por la restricción de $x \leq 5$, entonces $x = 5$ es el pico en esa dirección. Estos tres picos son los máximos locales (u *óptimos locales*) porque



■ **FIGURA 12.18**
Gráfica de ganancia de un ejemplo de programación no convexa.

■ **FIGURA 12.19**
Ejemplo de un problema de programación no convexa (que se muestra en la figura 12.18), donde Excel Solver obtiene tres soluciones diferentes a partir de tres soluciones iniciales distintas.

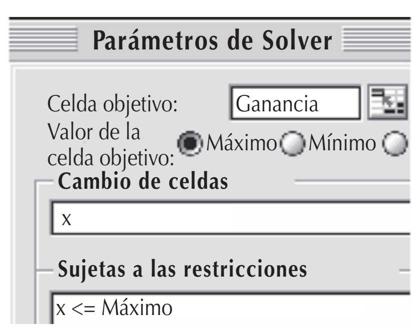
	A	B	C	D	E
1	**Solución de Solver**				
2	**(A partir de *x* = 0)**				
3					
4					Máximo
5		*x* =	0.371	?	5
6					
7	Ganancia =		$0.5x^5-6x^4+24.5x^3-39x^2+20x$		
8		=	\$3.19		

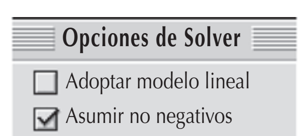
	A	B	C	D	E
1	**Solución de Solver**				
2	**(A partir de *x* = 3)**				
3					
4					Máximo
5		*x* =	3.126		5
6					
7	Ganancia =		$0.5x^5-6x^4+24.5x^3-39x^2+20x$		
8		=	\$6.13		

	A	B	C	D	E
1	**Solución de Solver**				
2	**(A partir de *x* = 4.7)**				
3					
4					Máximo
5		*x* =	5.000	?	5
6					
7	Ganancia=		$0.5x^5-6x^4+24.5x^3-39x^2+20x$		
8		=	\$0.00		

	B	C
7	Ganancia =	
8	=	=0.5*x^5–6*x^4+24.5*x^3–39*x^2+20*x

Nombre del rango	Celda
Máximo	E5
x	C5
Ganancia	C8





cada uno es un máximo de la gráfica dentro de una vecindad local de ese punto. Sin embargo, sólo el más grande de estos máximos locales es el *máximo global*, esto es, el punto más alto de toda la gráfica. Así, la hoja de cálculo de en medio en la figura 12.19 tuvo éxito al encontrar la solución óptima global en $x = 3.126$ con Ganancia = \$6.13.

El Excel Solver aplica el *método del gradiente reducido generalizado*, que adapta el método de búsqueda del gradiente descrito en la sección 12.5 para resolver problemas de programación convexa. Por tanto, este algoritmo se puede pensar como el procedimiento para escalar una montaña. Comienza con la solución inicial que se introdujo en las celdas cambiantes y después comienza a subir esa montaña hasta que llega al pico (o su ascenso se bloquea al llegar a una frontera impuesta por las restricciones). El procedimiento termina cuando alcanza este pico (o frontera) y reporta la solución. No tiene forma de detectar si existe alguna otra montaña más alta en algún lugar de la gráfica de ganancia.

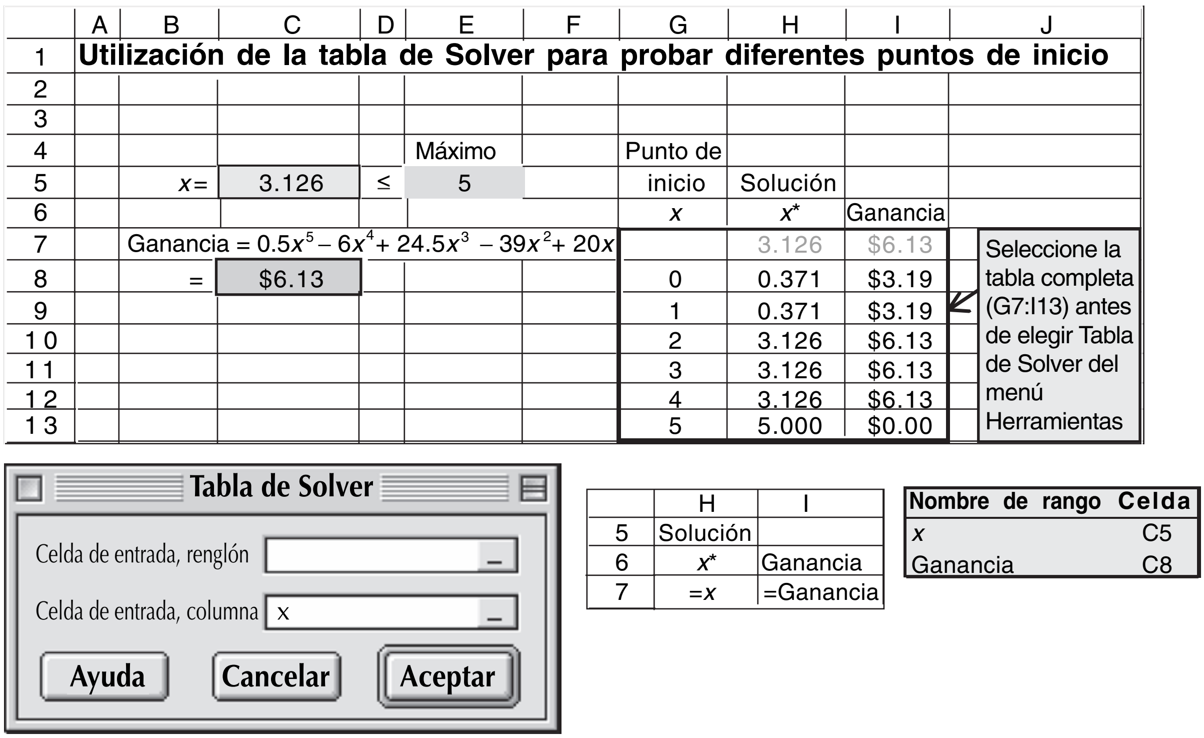
Lo mismo sucedería con cualquier otro procedimiento para escalar una montaña, como la SUMT (que se describió en la sección 12.9), la cual se detiene cuando encuentra un máximo local. Así, si la SUMT se aplicara a este ejemplo con cada una de las tres soluciones de prueba iniciales que se utilizaron en la figura 12.19, se encontrarían los mismos tres máximos locales que encontró el Excel Solver.

Un enfoque más sistemático para encontrar óptimos locales

Un enfoque usual de los problemas de programación no convexa "fáciles" es aplicar algún procedimiento algorítmico para escalar montañas que se detendrá cuando encuentre un *máximo local* y después reiniciará el algoritmo cierta cantidad de veces desde una variedad de soluciones de prueba iniciales (elegidas ya sea de manera aleatoria o como una sección cruzada sistemática) para encontrar tantos mínimos locales como sea posible. Después, el mejor de estos máximos locales se selecciona para su implantación. Por lo general, un procedimiento para escalar montañas se diseña para encontrar un máximo global cuando se cumplen todos los supuestos de la programación convexa, pero también puede operar para encontrar un máximo local cuando los supuestos no se cumplen.

Cuando se emplea el Solver de Excel, una forma sistemática de aplicar este enfoque es utilizar el complemento de la tabla de Solver que se proporciona en el OR Courseware. Para demostrar su uso se continuará con el modelo en hoja de cálculo que se presentó en la figura 12.19. Por otra parte, en la figura 12.20 se presenta cómo se utiliza la tabla de Solver para probar seis diferentes puntos de inicio (0, 1, 2, 3, 4 y 5) como las soluciones de prueba iniciales para este modelo al ejecutar los siguientes pasos. En el primer renglón de la tabla se introducen las fórmulas que se refieren a la celda cambiante, x (C5), y a la celda objetivo, Ganancia (C8). Los diferentes puntos de inicio se introducen en la primera columna de la tabla (G8:G13). Después se selecciona toda la tabla (G7:I13) y se elige Tabla de Solver de la pestaña Adiciones (en el caso de Excel 2007) o el menú Herramientas (para el caso de versiones anteriores de Excel). La celda de entrada que se especifica en el cuadro de diálogo de la tabla de Solver es la celda cambiante x (C5), puesto que es ahí donde se necesitan los diferentes puntos de inicio en la primera columna de la tabla que se introducirá. (En este cuadro de diálogo no se especifica el renglón de la celda cambiante porque sólo se utiliza una columna para definir los puntos de inicio.) Cuando se hace clic sobre Aceptar, el Solver de Excel resuelve de nuevo el problema para todos estos puntos de inicio en la primera columna y coloca los resultados correspondientes (el máximo local para x y la ganancia que produce en el primer renglón) en las otras columnas de la tabla.

El ejemplo tiene sólo una variable y por ende una sola celda cambiante. Sin embargo, la tabla de Solver también se puede utilizar para probar muchos puntos de inicio de problemas con dos variables (celdas cambiantes). Esto se hace utilizando el primer renglón y la primera columna de la tabla para especificar diferentes puntos de inicio para las dos celdas cambiantes. Se introduce una ecuación referida a la celda objetivo en la esquina superior izquierda de la tabla. Luego, se selecciona la tabla completa y se elige Tabla de Solver del menú Herramientas, con las dos celdas cambiantes seleccionadas como la celda de entrada columna y celda de entrada renglón. Luego, la tabla de Solver resuelve de nuevo el problema de cada combinación de puntos de inicio de las dos celdas cambiantes y llena el cuerpo de la tabla con el valor de la función objetivo que se encontró (un óptimo local) para cada una de estas combinaciones. (Vea la sección 6.8 para mayores detalles acerca de la preparación de la tabla de Solver en dos dimensiones.)



■ **FIGURA 12.20**
Aplicación de la tabla de Solver (un complemento que se proporciona en el OR Courseware) al ejemplo que se consideró en las figuras 12.18 y 12.19.

Para problemas con más de dos variables (celdas cambiantes), este mismo enfoque se puede seguir usando para probar puntos de inicio múltiples para cualquiera de las dos celdas cambiantes. Sin embargo, este enfoque se convierte en un modo muy incómodo de probar un amplio rango de puntos de inicio para todas las celdas cambiantes cuando hay más de tres o cuatro de estas celdas.

Desafortunadamente, es frecuente que no exista la garantía de encontrar una solución óptima global, sin importar cuántos puntos de inicio se prueben. También, si las gráficas de ganancia no son suaves (por ejemplo, si existen discontinuidades o espirales), entonces Solver puede no ser capaz siquiera de encontrar óptimos locales. Por fortuna, el *Premium Solver* de Excel proporciona otro procedimiento de búsqueda llamado *Evolutionary Solver* para intentar solucionar estos problemas de programación no convexa más complejos.

Evolutionary Solver

Frontline Systems, el desarrollador del Solver estándar que se incluye con Excel, ha desarrollado versiones Premium de Solver. Además, una versión de Premium Solver (Premium Solver for Education) está disponible en el OR Courseware (pero no se incluye con el Excel estándar). Todas las versiones de Premiun Solver, entre otras ésta, incluyen un procedimiento de búsqueda llamado **Evolutionary Solver** en el conjunto de herramientas disponible para buscar una solución óptima de un modelo. La filosofía del Evolutionary Solver se basa en la genética, evolución y supervivencia del más apto. Por tanto, este tipo de algoritmo algunas veces se conoce como **algoritmo genético**. La sección 13.4 está dedicada a describir cómo operan los algoritmos genéticos.

El Evolutionary Solver tiene tres ventajas cruciales sobre el Solver estándar (o cualquier otro algoritmo de programación convexa) para resolver problemas de programación no convexa. Primera, la complejidad de la función objetivo no lo afecta. Mientras la función se pueda evaluar para una solución de prueba dada, no importa si la función tiene rizos o discontinuidades o muchos óptimos

locales. Segunda, la complejidad de las restricciones dadas (se incluye restricciones no convexas) tampoco tienen un efecto sustancial en el Evolutionary Solver (aunque el *número* de restricciones sí lo hace). Tercera, como evalúa todas las poblaciones de soluciones de prueba, éstas no están necesariamente en la misma vecindad de la mejor solución de prueba actual. El Evolutionary Solver evita quedar atrapado en un óptimo local. En realidad, en el Evolutionary Solver está garantizado que finalmente se encontrará una solución óptima global para cualquier problema de programación no lineal (incluso problemas de programación no convexa), si éste se corre eternamente (lo que, por supuesto, no es práctico). Por lo tanto, el Evolutionary Solver está bien equipado para tratar con muchos problemas de programación no convexa relativamente pequeños.

Por otro lado, se debe puntualizar que el Evolutionary Solver no es una panacea. Primero, puede tomar *mucho* más tiempo que el Solver estándar para encontrar una solución final. Segundo, el Evolutionary no se desempeña bien con modelos que tienen muchas restricciones. Tercero, el Evolutionary Solver es un proceso aleatorio, por lo que, al correrlo de nuevo sobre un mismo modelo, es usual que se encuentre una solución final diferente. Finalmente, la mejor solución que se encuentre por lo general no es necesariamente la óptima (aunque puede estar muy cerca de serlo). El Evolutionary Solver no se mueve en forma continua hacia mejores soluciones. En lugar de ello es más un dispositivo de búsqueda inteligente, pues prueba distintas soluciones de prueba aleatorias. En consecuencia, mientras es muy probable que termine con una solución muy cercana a la óptima, casi nunca produce la solución óptima global exacta para la mayoría de los tipos de problemas de programación no lineal. En consecuencia, es frecuente obtener beneficios al correr el Solver estándar (GRG opción no lineal) después del Evolutionary Solver, a partir de la solución final que obtuvo este último, para ver si la solución se puede mejorar con una búsqueda alrededor de su vecindad.

12.11 CONCLUSIONES

En la práctica, es frecuente que los problemas de optimización incluyan un comportamiento *no lineal* que debe tomarse en cuenta. A veces es posible *reformular* las no linealidades para que se ajusten al formato de programación lineal, como en los problemas de *programación separable*. Sin embargo, muchas veces es necesario usar formulaciones de *programación no lineal*.

Al contrario del caso del método símplex para programación lineal, no existe un algoritmo eficiente que se pueda utilizar para resolver todos los problemas de programación no lineal. En realidad, algunos de estos problemas no se pueden resolver de modo satisfactorio por ningún método, pero se han hecho grandes progresos en ciertas clases importantes de problemas que incluyen *programación cuadrática*, *programación convexa* y algunos tipos especiales de *programación no convexa*. Se dispone de una gran variedad de algoritmos que casi siempre tienen un buen desempeño en estos casos. Algunos de estos algoritmos incorporan procedimientos de alta eficiencia para la *optimización no restringida* en una parte de cada iteración, y algunos emplean una sucesión de aproximaciones lineales o cuadráticas al problema original.

En los últimos años se ha manifestado un gran interés en el desarrollo de *paquetes de computadora* (software) confiables y de alta calidad para uso general en la aplicación del mejor de estos algoritmos. Por ejemplo, en el Systems Optimization Laboratory de la University of Stanford se han desarrollado varios paquetes poderosos, como el MINOS. Estos paquetes son de uso común en otros centros para la solución de problemas del tipo que se presentó en este capítulo (y de programación lineal). Las considerables mejoras que se han logrado, tanto en los algoritmos como en el software, permiten hoy que algunos problemas grandes estén dentro de la factibilidad computacional.

La investigación en el campo de la programación no lineal sigue muy activa.

REFERENCIAS SELECCIONADAS

1. Fiacco, A. V. y G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Classics in Applied Mathematics 4, Society for Industrial and Applied Mathematics, Philadelphia, 1990. (Reimpreso del libro clásico publicado en 1968.)
2. Fletcher, R., *Practical Methods of Optimization*, 2a. ed., Wiley, Nueva York, 2000.

3. Gill, P. E., W. Murray y M. H. Wright, *Practical Optimization*, Academic Press, Londres, 1981.
4. Hillier, F. S. y M. S. Hillier, *Introduction to Management Science: A Modeling and Case Studies Aproach with Spreadsheets*, 3a. ed., McGraw-Hill/Irwin, Burr Ridge, IL, 2008, capítulo 8.
5. Leyffer, S. y J. More (eds.), Special Issue on Deterministic Global Optimization and Applications, *Mathematical Programming*, Series B, **103**(2), junio de 2005.
6. Luenberger, D. y Y. Ye, *Linear and Nolinear Programming*, 3a. ed., Springer, Nueva York, 2008.
7. Miller, R. E., *Optimization: Foundations and Applications*, Wiley, Nueva York, 1999.
8. Rardin, D., *Optimization in Operations Research*, Prentice-Hall, Upper Saddle River, NJ, 1998.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos para el capítulo 12

Ejemplos de demostración incluidos en el OR Tutor:

Procedimiento de búsqueda del gradiente (*Gradient Search Procedure*)
Algoritmo de Frank-Wolfe (*Frank-Wolfe Algorithm*)
Técnica de minimización secuencial no restringida: SUMT (*Sequential Unconstrained Minimization Technique*)

Rutinas interactivas del IOR Tutorial:

Procedimiento interactivo de búsqueda en una dimensión (*Interactive One-Dimensional Search Procedure*)
Procedimiento de búsqueda del gradiente (*Interactive Gradient Search Procedure*)
Método símplex modificado (*Interactive Modified Simplex Method*)
Algoritmo de Frank-Wolfe (*Interactive Frank-Wolfe Algorithm*)

Rutinas automáticas del IOR Tutorial:

Procedimiento de búsqueda del gradiente (*Automatic Gradient Search Procedure*)
Técnica secuencial de minimización no restringida: SUMT (*Sequential Unconstrained Minimization Technique*)

Complemento de Excel:

Premium Solver for Education
Solver Table

Archivos para resolver los ejemplos, "Ch. 12—Nonlinear Programming"

Archivos de Excel
Archivo de LINGO/LINDO
Archivo de MPL/CPLEX/CONOPT/LGO

Glosario del capítulo 12

Vea en el apéndice 1 la documentación del software.

PROBLEMAS

Los símbolos a la izquierda de algunos problemas (o de sus incisos) significan lo siguiente:

D: El ejemplo de demostración correspondiente que se indicó en las Ayudas de aprendizaje puede ser útil.
I: Se sugiere usar la rutina interactiva indicada correspondiente que se acaba de listar (la impresión registra su trabajo).
C: Use la computadora con cualquier software (o según lo indique su profesor) para resolver el problema.

Un asterisco en el número del problema indica que al final del libro se da al menos una respuesta parcial.

12.1-1. Lea el artículo de referencia que describe en su totalidad el estudio de OR que se resume en el recuadro de aplicación que se presentó en la sección 12.1. Describa de manera concisa la forma en que se aplicó la programación lineal a este estudio. Después, elabore una lista de los beneficios financieros y no financieros que arrojó dicho estudio.

12.1-2. Considere el problema de la *mezcla de productos* que se presentó en el problema 3.1-11. Suponga que la empresa encuentra *elasticidad de precios* al vender sus tres productos, de tal suerte que las ganancias son distintas de las que se señalan en el capítulo 3. En particular, suponga que los costos unitarios respectivos de producir los artículos 1, 2 y 3 son de 25, 10 y 15 dólares, y que los precios que se requieren (en dólares) para poder vender x_1, x_2 y x_3 unidades son $(35 + 100x_1^{-\frac{1}{3}})$, $(15 + 40x_2^{-\frac{1}{4}})$ y $(20 + 50x_3^{-\frac{1}{2}})$, respectivamente.

Formule un modelo de programación no lineal del problema para determinar cuántas unidades de cada producto se deben producir para maximizar la ganancia.

12.1-3. En el problema de la P & T Co., que se describió en la sección 8.1, suponga que existe 10% de descuento en el precio del transporte para las cargas que *sobrepasen* las primeras 40 unidades en cada combinación de enlatadora y almacén. Dibuje figuras como las 12.3 y 12.4 que muestren los costos marginales del transporte de chícharos de la enlatadora 1 al almacén 1. Después, elabore el modelo completo de programación no lineal de este problema.

12.1-4. Un corredor de acciones, Richard Smith, acaba de recibir una llamada de su mejor cliente, Ann Hardy. Ella tiene 50 000 dólares para invertir y quiere usarlos para comprar dos tipos de acciones. La acción 1 es un bono sólido con un potencial de crecimiento respetable y poco riesgo. La acción 2 es más especulativa. Dos boletines de inversionistas afirman que tiene un gran potencial de crecimiento, pero la consideran riesgosa. Ann desea un rendimiento alto pero también tiene una elevada aversión al riesgo. En consecuencia, pide a Richard que analice qué mezcla de inversión de las dos acciones sería adecuada para ella.

Ann está acostumbrada a unidades de miles de dólares y bloques de 1 000 acciones. Con estas unidades, el precio por bloque de la acción l es de 20 y de 30 el de la 2. Después de investigar, Richard le presenta las siguientes estimaciones. El rendimiento esperado por bloque de la acción 1 es de 5 y de 10 el de la 2. La covarianza del rendimiento de un bloque de cada acción es de 5.

Sin asignar todavía un valor numérico específico al rendimiento mínimo aceptable, formule un modelo de programación no lineal del problema.

12.2-1. Reconsidere el problema 12.1-2. Verifique que se trata de un problema de programación convexa.

12.2-2. Reconsidere el problema 12.1-4. Demuestre que el modelo formulado es un problema de programación convexa con la prueba dada en el apéndice 2 para demostrar que la función objetivo que se debe minimizar es convexa.

12.2-3. Considere la variación del ejemplo de Wyndor Glass Co., que se representó en la figura 12.5, donde la segunda y tercera restricciones funcionales del problema original (sección 3.1) se sustituyeron por $9x_1^2 + 5x_2^2 \leq 216$. Demuestre que $(x_1, x_2) = (2, 6)$ con $Z = 36$ es óptima mostrando que la recta de la función objetivo, $36 = 3x_1 + 5x_2$ es *tangente* a esta frontera de restricción en (2, 6). (*Sugerencia:* Exprese x_2 en términos de x_1 sobre esta frontera y después obtenga la derivada respecto de x_1 para encontrar la pendiente de la frontera.)

12.2-4. Considere la variación del problema de Wyndor Glass Co., que se presentó en la figura 12.6, donde la función objetivo original (vea la sección 3.1) se sustituyó por $Z = 126x_1 - 9x_1^2 + 182x_2 - 13x_2^2$. Demuestre que $(x_1, x_2) = (\frac{8}{3}, 5)$ con $Z = 857$ es óptima demostrando que la elipse $857 = 126x_1 - 9x_1^2 + 182x_2 - 13x_2^2$ es tangente a la frontera de la restricción $3x_1 + 2x_2 = 18$ en $(\frac{8}{3}, 5)$. (*Sugerencia:* Obtenga x_2 en términos de x_1 en la elipse y después derive la expresión respecto de x_1 para encontrar la pendiente de la elipse.)

12.2-5. Considere la siguiente función:

$$f(x) = 240x - 300x^2 + 10x^3.$$

a) Utilice la primera y segunda derivadas para encontrar los máximos y mínimos locales de $f(x)$.
b) Utilice la primera y segunda derivadas para demostrar que $f(x)$ no tiene ni máximo ni mínimo globales porque es no acotada en ambas direcciones.

12.2-6. Demuestre si cada una de las siguientes funciones, es convexa, cóncava o ninguna de las dos.

a) $f(x) = 10x - x^2$
b) $f(x) = x^4 + 6x^2 + 12x$
c) $f(x) = 2x^3 - 3x^2$
d) $f(x) = x^4 + x^2$
e) $f(x) = x^3 + x^4$

12.2-7.* Utilice la prueba del apéndice 2 para determinar si cada una de las siguientes funciones es convexa, cóncava o ninguna de las dos.

a) $f(\mathbf{x}) = x_1x_2 - x_1^2 - x_2^2$
b) $f(\mathbf{x}) = 3x_1 + 2x_1^2 + 4x_2 + x_2^2 - 2x_1x_2$
c) $f(\mathbf{x}) = x_1^2 + 3x_1x_2 + 2x_2^2$
d) $f(\mathbf{x}) = 20x_1 + 10x_2$
e) $f(\mathbf{x}) = x_1x_2$

12.2-8. Considere la siguiente función:

$$f(\mathbf{x}) = 5x_1 + 2x_2^2 + x_3^2 - 3x_3x_4 + 4x_4^2 + 2x_5^4 + x_5^2 + 3x_5x_6 + 6x_6^2 + 3x_6x_7 + x_7^2.$$

Demuestre que $f(\mathbf{x})$ es convexa; exprésela como suma de funciones de una o dos variables y compruebe (vea el apéndice 2) que todas estas funciones son convexas.

12.2-9. Considere el siguiente problema de programación no lineal:

Maximizar $f(\mathbf{x}) = x_1 + x_2,$

sujeta a

$x_1^2 + x_2^2 \leq 1$

y

$x_1 \geq 0, \quad x_2 \geq 0.$

a) Verifique que se trata un problema de programación convexa.
b) Resuelva el problema gráficamente.

12.2-10. Considere el siguiente problema de programación no lineal:

Minimizar $Z = x_1^4 + 2x_2^2,$

sujeta a

$x_1^2 + x_2^2 \geq 2.$

(Sin restricciones de no negatividad.)

a) Use el análisis genérico para determinar si la región factible es un conjunto convexo.
b) Ahora utilice álgebra y cálculo para determinar si la región factible es un conjunto convexo.

12.3-1. Reconsidere el problema 12.1-3. Demuestre que se trata de un problema de programación no convexa.

12.3-2. Considere el siguiente problema de optimización restringida:

Maximizar $f(x) = -120x + 15x^2 - 10x^3,$

sujeta a

$x \geq 0.$

Utilice sólo la primera y segunda derivadas de $f(x)$ para obtener una solución óptima.

12.3-3. Considere el siguiente problema de programación no lineal:

Minimizar $Z = x_1^4 + 2x_1^2 + 2x_1x_2 + 4x_2^2,$

sujeta a

$2x_1 + x_2 \geq 10$

$x_1 + 2x_2 \geq 10$

y

$x_1 \geq 0, \quad x_2 \geq 0.$

a) De los tipos especiales de problemas de programación no lineal que se describieron en la sección 12.3, ¿a cuál o cuáles se puede ajustar este problema? Justifique su respuesta.
b) Ahora suponga que el problema cambia un poco al sustituir las restricciones de no negatividad por $x_1 \geq 1$ y $x_2 \geq 1$. Convierta este nuevo problema en uno equivalente que tenga sólo dos restricciones funcionales, dos variables y dos restricciones de no negatividad.

12.3-4. Considere el siguiente problema de programación geométrica:

Minimizar $f(\mathbf{x}) = 2x_1^{-2}x_2^{-1} + x_2^{-2},$

sujeta a

$4x_1x_2 + x_1^2x_2^2 \leq 12$

y

$x_1 \geq 0, \quad x_2 \geq 0.$

a) Transforme este problema en uno equivalente de programación convexa.
b) Utilice la prueba que se presentó en el apéndice 2 para verificar que el modelo reformulado en el inciso *a*) es un problema de programación convexa.

12.3-5. Considere el siguiente problema de programación lineal fraccional:

Maximizar $f(\mathbf{x}) = \dfrac{10x_1 + 20x_2 + 10}{3x_1 + 4x_2 + 20},$

sujeta a

$x_1 + 3x_2 \leq 50$
$3x_1 + 2x_2 \leq 80$

y

$x_1 \geq 0, \quad x_2 \geq 0.$

a) Transforme este problema en uno equivalente de programación lineal.
C ***b)*** Utilice la computadora para resolver el modelo formulado en el inciso *a*). ¿Cuál es la solución óptima que resulta para el problema original?

12.3-6. Considere las expresiones en notación matricial que se presentaron en la sección 12.7 para la forma general de las condiciones KKT del problema de programación cuadrática. Demuestre que el problema de encontrar una solución factible para estas condiciones es un problema de complementariedad lineal, como el que se presentó en la sección 12.3; para ello, identifique **w**, **z**, **q** y **M** en términos de los vectores y matrices de la sección 12.7.

12.4-1.* Considere el siguiente problema:

Maximizar $f(x) = x^3 + 2x - 2x^2 - 0.25x^4.$

I ***a)*** Aplique el método de bisección para resolver este problema de manera aproximada. Use una tolerancia de error $\epsilon = 0.04$ y cotas iniciales $\underline{x} = 0, \bar{x} = 2.4$.
b) Aplique el método de Newton, con $\epsilon = 0.001$ y $x_1 = 1.2$, en este problema.

I **12.4-2.** Utilice el método de bisección con una tolerancia de error $\epsilon = 0.04$ y las cotas iniciales dadas para obtener de forma interactiva una solución (aproximada) de los problemas:

a) Maximizar $f(x) = 6x - x^2$, con $\underline{x} = 0, \bar{x} = 4.8.$
b) Minimizar $f(x) = 6x + 7x^2 + 4x^3 + x^4$, con $\underline{x} = -4, \bar{x} = 1.$

12.4-3. Considere el siguiente problema:

Maximizar $f(x) = 48x^5 + 42x^3 + 3.5x - 16x^6 - 61x^4 - 16.5x^2.$

I *a*) Aplique el método de bisección para encontrar una solución (aproximada) a este problema. Emplee una tolerancia de error $\epsilon = 0.08$ y cotas iniciales $\underline{x} = -1, \bar{x} = 4$.

b) Aplique el método de Newton a este problema con $\epsilon = 0.001$ y $x_1 = 1$.

12.4-4. Considere el siguiente problema:

Maximizar $\quad f(x) = 10x^3 + 60x - 2x^6 - 3x^4 - 12x^2.$

I *a*) Aplique el método de bisección para encontrar una solución (aproximada) a este problema. Emplee una tolerancia de error $\epsilon = 0.07$ y determine las cotas iniciales apropiadas mediante inspección.

b) Aplique el método de Newton a este problema con $\epsilon = 0.001$ y $x_1 = 1$.

12.4-5. Considere el siguiente problema de programación convexa:

Minimizar $\quad Z = x^4 + x^2 - 4x,$

sujeta a

$x \leq 2 \quad \text{y} \quad x \geq 0.$

a) Haga un cálculo sencillo sólo para verificar si la solución óptima está dentro del intervalo $0 \leq x \leq 1$ o de $1 \leq x \leq 2$. (*No* obtenga la solución óptima para determinar en qué intervalo debe estar.) Explique su lógica.

I *b*) Use el método de bisección con cotas iniciales $\underline{x} = 0, \bar{x} = 2$ y con una tolerancia de error $\epsilon = 0.02$ para obtener, de forma interactiva, una solución (aproximada) de este problema.

c) Aplique el método de Newton a este problema con $\epsilon = 0.0001$ y $x_1 = 1$.

12.4-6. Considere el problema de maximizar una función diferencial $f(x)$ con una sola variable no restringida x. Sean $\underline{x}_0$ y $\bar{x}_0$, las respectivas cotas inferior y superior válidas sobre el mismo máximo global (si existe). Pruebe las siguientes propiedades generales del método de bisección (como se hizo en la sección 12.4) para intentar resolver un problema de este tipo.

a) Dadas $\underline{x}_0, \bar{x}_0$ y $\epsilon = 0$, la sucesión de soluciones de prueba seleccionadas por la *regla del punto medio* debe *converger* hacia una solución límite. [*Sugerencia:* Demuestre que $\lim_{n\to\infty}(\bar{x}_n - \underline{x}_n) = 0$, donde $\bar{x}_n$ y $\underline{x}_n$ son las cotas superior e inferior identificadas en la iteración n.]

b) Si $f(x)$ es cóncava [de manera que $df(x)/dx$ es una función monótona decreciente de x], entonces la solución límite del inciso *a*) debe ser un máximo global.

c) Si $f(x)$ no es cóncava en todas sus partes, pero sería cóncava si su dominio se restringiera al intervalo entre $\underline{x}_0$ y $\bar{x}_0$, entonces la solución límite del inciso *a*) debe ser un máximo global.

d) Si $f(x)$ no es cóncava ni siquiera en el intervalo entre $\underline{x}_0$ y $\bar{x}_0$, entonces la solución límite del inciso *a*) no necesariamente es un máximo global. (Pruebe esto con la construcción de un contraejemplo gráfico.)

e) Si $df(x)/dx < 0$ para toda x, entonces no existe $\underline{x}_0$. Si $df(x)/dx > 0$ para toda x, entonces no existe $\bar{x}_0$. En cualquier caso, $f(x)$ no tiene un máximo global.

f) Si $f(x)$ es cóncava y $\lim_{x\to-\infty} df(x)/dx < 0$, entonces no existe $\underline{x}_0$. Si $f(x)$ es cóncava y $\lim_{x\to\infty} df(x)/dx > 0$, entonces no existe $\bar{x}_0$. En cualquier caso, $f(x)$ no tiene un máximo global.

I **12.4-7.** Considere el siguiente problema de programación convexa linealmente restringida:

Maximizar $\quad f(\mathbf{x}) = 32x_1 + 50x_2 - 10x_2^2 + x_2^3 - x_1^4 - x_2^4,$

sujeta a

$$3x_1 + x_2 \leq 11$$
$$2x_1 + 5x_2 \leq 16$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

Pase por alto las restricciones y resuelva los problemas de *optimización restringida de una variable*. Utilice cálculo para resolver el problema que incluye a x_1 y use el método de bisección con $\epsilon = 0.001$ y cotas iniciales de 0 y 4, para resolver el problema que incluye a x_2. Demuestre que la solución que obtenga para (x_1, x_2) satisface todas las restricciones, de manera que en realidad es óptima para el problema original.

12.5-1. Considere el siguiente problema de optimización no restringida:

Maximizar $\quad f(\mathbf{x}) = 2x_1x_2 + x_2 - x_1^2 - 2x_2^2.$

D,I *a*) Si se inicia con la solución de prueba inicial $(x_1, x_2) = (1, 1)$, aplique de manera interactiva el procedimiento de búsqueda del gradiente con $\epsilon = 0.25$ para obtener una solución aproximada.

b) Resuelva el sistema de ecuaciones lineales que obtenga con $\nabla f(\mathbf{x}) = \mathbf{0}$ para llegar a la solución exacta.

c) Con referencia a la figura 12.14 como ejemplo de un problema similar, dibuje la trayectoria de las soluciones de prueba del inciso *a*). Después muestre la *continuación* aparente de esta trayectoria con la mejor estimación de las tres soluciones de prueba que siguen [según el patrón del inciso *a*) y la figura 12.14]. También muestre la solución exacta del inciso *b*) hacia la que converge esta sucesión de soluciones de prueba.

C *d*) Aplique a este problema la rutina automática del procedimiento de búsqueda del gradiente en el IOR Tutorial (con $\epsilon = 0.01$).

D,I,C **12.5-2.** A partir de la solución de prueba inicial $(x_1, x_2) = (1, 1)$, haga dos iteraciones interactivas del procedimiento de búsqueda del gradiente para comenzar la solución del siguiente problema y aplique la rutina automática para desarrollar este procedimiento (con $\epsilon = 0.01$).

Maximizar $\quad f(\mathbf{x}) = 60x_1x_2 - 15x_1^2 - 80x_2^2.$

Después resuelva $\nabla f(\mathbf{x}) = \mathbf{0}$ directamente para obtener la solución exacta.

D,I,C **12.5-3.*** A partir de la solución de prueba inicial $(x_1, x_2) = (0, 0)$, aplique el procedimiento interactivo de búsqueda del gradiente con $\epsilon = 0.3$ para obtener una solución aproximada del siguiente problema y después la rutina automática para desarrollar este procedimiento (con $\epsilon = 0.01$).

Maximizar $\quad f(\mathbf{x}) = 8x_1 - x_1^2 - 12x_2 - 2x_2^2 + 2x_1x_2.$

Después resuelva $\nabla f(\mathbf{x}) = \mathbf{0}$ directamente para obtener la solución exacta.

D,I,C **12.5-4.** A partir de la solución de prueba inicial $(x_1, x_2) = (0, 0)$, aplique dos iteraciones interactivas del procedimiento de búsqueda del gradiente y después la rutina automática para desarrollar este procedimiento (con $\epsilon = 0.01$).

$$\text{Maximizar} \qquad f(\mathbf{x}) = 6x_1 + 2x_1x_2 - 2x_2 - 2x_1^2 - x_2^2.$$

Después resuelva $\nabla f(\mathbf{x}) = \mathbf{0}$ directamente para obtener la solución exacta.

12.5-5. A partir de la solución de prueba inicial $(x_1, x_2) = (0, 0)$ aplique a mano una iteración del procedimiento de búsqueda del gradiente al siguiente problema:

$$\text{Maximizar} \qquad f(\mathbf{x}) = 4x_1 + 2x_2 + x_1^2 - x_1^4 - 2x_1x_2 - x_2^2.$$

Para completar esta iteración, obtenga una aproximación de t^* mediante la aplicación de *dos* iteraciones del método de bisección con cotas iniciales $\underline{t} = 0, \bar{t} = 1$.

12.5-6. Considere el siguiente problema de optimización no restringida:

$$\text{Maximizar} \qquad f(\mathbf{x}) = 3x_1x_2 + 3x_2x_3 - x_1^2 - 6x_2^2 - x_3^2.$$

a) Describa cómo se puede reducir la solución de este problema a la solución de uno de optimización no restringida con *dos variables*.

D,I ***b)*** A partir de la solución de prueba inicial $(x_1, x_2, x_3) = (1, 1, 1)$, aplique el procedimiento interactivo de búsqueda del gradiente con $\epsilon = 0.05$ para obtener una solución (aproximada) del problema con dos variables identificado en el inciso *a*).

C ***c)*** Repita el inciso *b*) con la rutina automática para desarrollar este procedimiento (con $\epsilon = 0.005$).

D,I,C **12.5-7.*** A partir de la solución de prueba inicial $(x_1, x_2) = (0, 0)$, aplique el *procedimiento interactivo de búsqueda del gradiente* con $\epsilon = 1$ para obtener una solución (aproximada) para el siguiente problema y luego aplique la rutina automática para desarrollar este procedimiento (con $\epsilon = 0.01$).

$$\text{Maximizar} \qquad f(\mathbf{x}) = x_1x_2 + 3x_2 - x_1^2 - x_2^2.$$

12.6-1. Reconsidere el modelo de programación convexa de una variable que se plantea en el problema 12.4-5. Utilice las condiciones KKT para derivar una solución óptima.

12.6-2. Reconsidere el problema 12.2-9. Use las condiciones KKT para verificar si $(x_1, x_2) = (1/\sqrt{2}, 1/\sqrt{2})$ es óptima.

12.6-3.* Reconsidere el modelo del problema 12.3-3. ¿Cuáles son las condiciones de este modelo? Úselas para determinar si $(x_1, x_2) = (0, 10)$ puede ser óptima.

12.6-4. Considere el siguiente problema de programación convexa.

$$\text{Maximizar} \qquad f(\mathbf{x}) = 12x_1 - x_1^2 + 50x_2 - x_2^2,$$

sujeta a

$$x_1 \leq 10,$$
$$x_2 \leq 15,$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

a) Utilice las condiciones KKT de este problema para derivar una solución óptima.

b) Descomponga este problema en dos problemas de optimización restringida separados que incluyan sólo x_1 y sólo x_2. Grafique la función objetivo de cada uno sobre la región factible para *demostrar* que el valor de x_1 o de x_2 que obtuvo en el inciso *a*) es óptimo. Después *pruebe* que este valor es óptimo si se usa nada más la primera y segunda derivadas de la función objetivo y las restricciones de los respectivos problemas.

12.6-5. Considere el siguiente problema de optimización linealmente restringida:

$$\text{Maximizar} \qquad f(\mathbf{x}) = \ln(x_1 + 1) - x_2^2,$$

sujeta a

$$x_1 + 2x_2 \leq 3$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0,$$

donde ln denota el logaritmo natural.

a) Verifique que se trata de un problema de programación convexa.

b) Utilice las condiciones KKT para derivar una solución óptima.

c) Utilice un razonamiento intuitivo para demostrar que la solución obtenida en *b*) en realidad es óptima.

12.6-6.* Considere el problema de programación no lineal que se presentó con el número 10.3-11. Determine si $(x_1, x_2) = (1, 2)$ puede ser óptima si se aplican las condiciones KKT.

12.6-7. Considere el siguiente problema de programación no lineal:

$$\text{Maximizar} \qquad f(\mathbf{x}) = \frac{x_1}{x_2 + 1},$$

sujeta a

$$x_1 - x_2 \leq 2$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

a) Use las condiciones KKT para demostrar que $(x_1, x_2) = (4, 2)$ *no* es óptima.

b) Derive una solución que cumpla las condiciones KKT.

c) Demuestre que este problema no es de programación convexa.

d) A pesar de la conclusión del inciso *c*), utilice un razonamiento *intuitivo* para demostrar que la solución que se obtuvo en el inciso *b*) es óptima. [La razón teórica es que $f(\mathbf{x})$ es *seudocóncava*.]

e) Use el hecho de que éste es un problema de programación lineal fraccional para transformarlo en uno equivalente de programación lineal. Resuelva el último e identifique la solución óptima del problema original. (*Sugerencia:* Use la restricción de igualdad en el problema de programación lineal para sustituir una variable del modelo y luego resuelva este modelo en forma gráfica.)

12.6-8.* Use las condiciones KKT para derivar una solución óptima en cada uno de los siguientes problemas.

a) Maximizar $\qquad f(\mathbf{x}) = x_1 + 2x_2 - x_2^3,$

sujeta a

$x_1 + x_2 \leq 1$

y

$x_1 \geq 0, \quad x_2 \geq 0.$

***b*)** Maximizar $\quad f(\mathbf{x}) = 20x_1 + 10x_2,$

sujeta a

$x_1^2 + x_2^2 \leq 1$
$x_1 + 2x_2 \leq 2$

y

$x_1 \geq 0, \quad x_2 \geq 0.$

12.6-9. ¿Cuáles son las condiciones KKT aplicables a problemas de programación no lineal de la forma siguiente?

Minimizar $\quad f(\mathbf{x}),$

sujeta a

$g_i(\mathbf{x}) \geq b_i, \qquad$ para $i = 1, 2, \ldots, m$

y

$\boldsymbol{x} \geq \mathbf{0}.$

(*Sugerencia:* Convierta esta forma a nuestra forma estándar, que en este capítulo quedó implícita, mediante las técnicas que se presentaron en la sección 4.6; después, aplique las condiciones KKT como se dan en la sección 12.6.)

12.6-10. Considere el siguiente problema de programación no lineal:

Minimizar $\quad Z = 2x_1^2 + x_2^2,$

sujeta a

$x_1 + x_2 = 10$

y

$x_1 \geq 0, \quad x_2 \geq 0.$

***a*)** De los tipos especiales de problemas de programación no lineal descritos en la sección 12.3, ¿a cuál o a cuáles se puede ajustar este problema? Justifique su respuesta. (*Sugerencia:* Primero convierta este problema en uno equivalente de programación no lineal que se ajuste a la forma dada en el segundo párrafo del capítulo, con $m = 2$ y $n = 2$.)
***b*)** Obtenga las condiciones KKT aplicables a este problema.
***c*)** Use las condiciones KKT para derivar una solución óptima.

12.6-11. Considere el siguiente problema de programación linealmente restringida:

Minimizar $\quad f(\mathbf{x}) = x_1^3 + 4x_2^2 + 16x_3,$

sujeta a

$x_1 + x_2 + x_3 = 5$

y

$x_1 \geq 1, \quad x_2 \geq 1, \quad x_3 \geq 1.$

***a*)** Convierta este problema en uno de programación no lineal equivalente que se ajuste a la forma que se dio al principio del capítulo (segundo párrafo), con $m = 2$ y $n = 3$.
***b*)** Utilice la forma que obtuvo en el inciso *a*) para construir las condiciones KKT de este problema.
***c*)** Utilice las condiciones KKT para verificar si $(x_1, x_2, x_3) = (2, 1, 2)$ es óptima.

12.6-12. Considere el siguiente problema de programación convexa linealmente restringida:

Minimizar $\quad Z = x_1^2 - 6x_1 + x_2^3 - 3x_2,$

sujeta a

$x_1 + x_2 \leq 1$

y

$x_1 \geq 0, \quad x_2 \geq 0.$

***a*)** Obtenga las condiciones KKT de este problema.
***b*)** Utilice las condiciones KKT para verificar si $(x_1, x_2) = (\frac{1}{2}, \frac{1}{2})$ es una solución óptima.
***c*)** Utilice las condiciones KKT para derivar una solución óptima.

12.6-13. Considere el siguiente problema de programación convexa linealmente restringida:

Maximizar $\quad f(\mathbf{x}) = 8x_1 - x_1^2 + 2x_2 + x_3,$

sujeta a

$x_1 + 3x_2 + 2x_3 \leq 12$

y

$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0.$

***a*)** Use las condiciones KKT para demostrar que $(x_1, x_2, x_3) = (2, 2, 2)$ *no* es una solución óptima.
***b*)** Utilice las condiciones KKT para derivar una solución óptima. (*Sugerencia:* Haga un análisis intuitivo preliminar para determinar el caso más prometedor sobre las variables que son diferentes de cero y las que son iguales a cero.)

12.6-14. Utilice las condiciones KKT para determinar si $(x_1, x_2, x_3) = (1, 1, 1)$ puede ser óptima para el problema siguiente:

Minimizar $\quad Z = 2x_1 + x_2^3 + x_3^2,$

sujeta a

$x_1^2 + 2x_2^2 + x_3^2 \geq 4$

y

$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0.$

12.6-15. Reconsidere el modelo que se presentó en el problema 12.2-10. ¿Cuáles son las condiciones KKT de este problema? Utilice estas condiciones para determinar si $(x_1, x_2) = (1, 1)$ puede ser óptima.

12.6-16. Reconsidere el modelo de programación convexa linealmente restringida del problema 12.4-7. Use las condiciones KKT para determinar si $(x_1, x_2) = (2, 2)$ puede ser óptima.

12.7-1. Considere el problema de programación cuadrática que se presentó en la sección 12.7.

a) Utilice la prueba que se dio en el apéndice 2 para demostrar que la función objetivo es *estrictamente cóncava*.
b) Verifique que la función objetivo es estrictamente cóncava al demostrar que **Q** es una matriz *positiva definida*; esto es, $\mathbf{x}^T\mathbf{Q}\mathbf{x} > \mathbf{0}$ para toda $\mathbf{x} \neq \mathbf{0}$. (*Sugerencia:* Reduzca $\mathbf{x}^T\mathbf{Q}\mathbf{x}$ a una suma de cuadrados.)
c) Demuestre que $x_1 = 12$, $x_2 = 9$ y $u_1 = 3$ satisfacen las condiciones KKT cuando se escriben en la forma que se dio en la sección 12.6.

12.7-2.* Considere el siguiente problema de programación cuadrática.

$$\text{Maximizar} \quad f(\mathbf{x}) = 8x_1 - x_1^2 + 4x_2 - x_2^2,$$

sujeta a

$$x_1 + x_2 \leq 2$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

a) Utilice las condiciones KKT para derivar una solución óptima.
b) Ahora suponga que este problema debe resolverse por el método símplex modificado. Formule el problema de programación lineal que se va a manejar en forma explícita y después identifique la restricción de complementariedad adicional a la que obliga el algoritmo de manera automática.
I **c**) Aplique el método símplex modificado al problema que se formuló en el inciso *b*).
C **d**) Utilice la computadora para resolver el problema de programación cuadrática en forma directa.

12.7-3. Considere el siguiente problema de programación cuadrática.

$$\text{Maximizar} \quad f(\mathbf{x}) = 250x_1 - 25x_1^2 + 100x_2 - 100x_2^2 + 90x_1x_2,$$

sujeta a

$$20x_1 + 5x_2 \leq 90$$
$$10x_1 + 10x_2 \leq 60$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

Suponga que este problema se debe resolver por el método símplex modificado.
a) Formule el problema de programación lineal que se debe manejar en forma explícita; después identifique la restricción de complementariedad adicional a la que obliga automáticamente el algoritmo.
I ***b***) Aplique el método símplex modificado al problema que se formuló en el inciso *a*).

12.7-4. Considere el siguiente problema de programación cuadrática.

$$\text{Maximizar} \quad f(\mathbf{x}) = 2x_1 + 3x_2 - x_1^2 - x_2^2,$$

sujeta a

$$x_1 + x_2 \leq 2$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

a) Utilice las condiciones KKT para derivar una solución óptima de manera directa.
b) Ahora suponga que este problema debe resolverse por el método símplex modificado. Formule el problema de programación lineal que se debe manejar en forma explícita y después identifique la restricción de complementariedad adicional a la que obliga el algoritmo.
c) Sin aplicar el método símplex modificado, demuestre que la solución que obtuvo en el inciso *a*) es óptima ($Z = 0$) para el problema equivalente que se formuló en el inciso *b*).
I ***d***) Aplique el método símplex modificado al problema que se presentó en el inciso *b*).
C ***e***) Use la computadora para resolver el problema de programación cuadrática.

12.7-5. Reconsidere la primera variación de programación cuadrática del problema de la Wyndor Glass Co., que se presentó en la sección 12.2 (vea la figura 12.6). Analice este problema de acuerdo con las instrucciones de los incisos *a*), *b*) y *c*) del problema 12.7-4.

C **12.7-6.** Reconsidere el problema 12.1-4 y su modelo de programación cuadrática.
a) Presente este modelo [y los valores de $R(\mathbf{x})$ y $V(\mathbf{x})$] en una hoja de cálculo de Excel.
b) Resuelva cuatro casos de este modelo: rendimiento esperado mínimo aceptable = 13, 14, 15, 16.
c) Para distribuciones de probabilidad típicas (con media μ y varianza σ^2) del rendimiento total de una carrera, la probabilidad de que el rendimiento exceda $\mu - \sigma$ es alta (de 0.8 a 0.9) y la probabilidad de que exceda $\mu - 3\sigma$ es muy alta (muy cercana a 0.999). Calcule $\mu - \sigma$ y $\mu - 3\sigma$ de los cuatro portafolios que obtuvo en el inciso ***b***). ¿Cuál cartera dará la μ más alta entre las que también cumplen que $\mu - \sigma \geq 0$?

12.8-1. La corporación MFG planea fabricar y vender tres productos. Sea x_1, x_2 y x_3 el número de unidades de cada producto que debe fabricarse. Las estimaciones preliminares de la ganancia potencial son las siguientes.

La ganancia unitaria de las primeras 15 unidades fabricadas del producto 1 será de aproximadamente 500 dólares. La ganancia unitaria de cualquier unidad adicional será de sólo 60 dólares. Se estima que la ganancia unitaria de las primeras 20 unidades del producto 2 será de 400 dólares, las siguientes 20 unidades tendrán una ganancia de 200 dólares y será de 100 dólares la de cualquier unidad adicional. La ganancia unitaria de las primeras 20 unidades del producto 3 será de 600 dólares, la de las siguientes diez unidades disminuirá a 400 dólares y a 200 dólares la de cualquier unidad adicional.

Ciertas limitaciones en el uso de los recursos necesarios imponen las siguientes restricciones de fabricación:

$$\begin{aligned} 2x_1 + 3x_2 + 4x_3 &\leq 180 \\ 3x_1 + x_2 &\leq 150 \\ x_1 + 3x_3 &\leq 100 \end{aligned}$$

La administración desea saber qué valores de x_1, x_2 y x_3 deben elegirse a fin de maximizar la ganancia total.
a) Grafique la ganancia de cada producto.

b) Use programación separable para formular un modelo de programación lineal para representar el problema.

C *c*) Resuelva el modelo. ¿Qué se recomienda a la administración sobre los valores de x_1, x_2 y x_3?

d) Ahora suponga que existe una restricción adicional sobre la ganancia de los productos 1 y 2, que debe dar un total de 20 000 dólares. Utilice la técnica que se presentó en la subsección de "Extensiones" de la sección 12.8 para agregar esta restricción al modelo que se formuló en el inciso *b*).

C *e*) Repita el inciso *c*) para el modelo que se formuló en *d*).

12.8-2.* La Dorwyn Company tiene dos productos nuevos para competir con los dos productos nuevos de la Wyndor Glass (sección 3.1). Con unidades de cientos de dólares en la función objetivo, se formuló el siguiente modelo de programación lineal para determinar la mezcla de productos más redituable.

Maximizar $Z = 4x_1 + 6x_2$,

sujeta a

$$x_1 + 3x_2 \leq 8$$
$$5x_1 + 2x_2 \leq 14$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

Sin embargo, debido a la fuerte competencia de la Wyndor, la administración de Dorwyn sabe que debe hacer un gran esfuerzo para generar grandes ventas de sus productos. En particular, estima que lograr una tasa de producción y ventas de x_1 unidades del producto 1 por semana requerirá un costo semanal de comercialización de x_1^3 cientos de dólares. Los costos de comercialización correspondientes del producto 2 se estiman en $2x_2^2$ cientos de dólares. En consecuencia, la función objetivo del modelo debe ser $Z = 4x_1 + 6x_2 - x_1^3 - 2x_2^2$.

La administración de Dorwyn desea usar el modelo revisado para determinar la mezcla más redituable.

a) Verifique que $(x_1, x_2) = (2/\sqrt{3}, \frac{3}{2})$ es una solución óptima aplicando las condiciones KKT.

b) Construya tablas con los datos de ganancias de cada producto con tasas de producción de 0, 1, 2, 3.

c) Dibuje una figura como la 12.15*b* con los puntos de ganancia semanales de cada producto cuando la tasa de producción es 0, 1, 2, 3. Conecte los pares de puntos consecutivos con líneas punteadas.

d) Con programación separable y la figura formule un modelo aproximado de programación lineal.

C *e*) Resuelva el modelo. ¿Qué le indica el resultado a la administración de Dorwyn sobre la mezcla de productos?

12.8-3. La compañía J. B. Jensen se especializa en la producción de sierras eléctricas para uso doméstico. Las ventas son estables durante el año y suben durante la temporada de Navidad. Como el trabajo de producción requiere de mucha experiencia, la compañía mantiene un nivel de empleados estable y usa tiempo extra para aumentar la producción en noviembre. Los trabajadores aprecian esta oportunidad de ganar dinero extra para las fiestas.

B. J. Jensen, Jr., el presidente actual, estudia los planes de producción para el próximo noviembre y cuenta con los siguientes datos.

	Producción mensual máxima*		Utilidad por unidad producida	
	Tiempo normal	**Horas extra**	**Tiempo normal**	**Horas extras**
Sierras	12 000	8 000	\$240	\$80
Taladros	20 000	12 000	\$160	\$120

*** Se supone que los suministros de materiales de los proveedores son adecuados.**

Sin embargo, el Sr. Jensen ahora sabe que, además del número limitado de horas de mano de obra, otros dos factores limitarán los niveles de producción en noviembre. Uno es que los proveedores de los motores sólo podrán entregar 40 000 unidades para noviembre (8 000 más que la entrega usual). Cada sierra y cada taladro requieren una unidad. Segundo, el proveedor de una parte clave de los ensambles de engranes podrá surtir 60 000 para noviembre (16 000 más que otros meses). Cada sierra requiere dos de estas partes, y cada taladro, una.

El Sr. Jensen desea determinar cuántas sierras y taladros eléctricos debe producir en noviembre para maximizar la ganancia total de la compañía.

a) Grafique la ganancia de cada producto.

b) Use programación separable para formular un modelo de programación lineal para representar el problema.

C *c*) Resuelva el modelo. ¿Qué indica la solución sobre cuántas sierras y taladros eléctricos se deben producir en noviembre?

12.8-4. Reconsidere el modelo de programación convexa linealmente restringida del problema 12.4-7.

a) Use la técnica de programación separable de la sección 12.8 para formular un modelo de programación lineal aproximado. Use $x_1 = 0, 1, 2, 3$ y $x_2 = 0, 1, 2, 3$ como puntos de salto de las funciones lineales por partes.

C *b*) Utilice el método símplex para resolver el modelo formulado en *a*). Después reescriba esta solución en términos de las variables *originales* del problema.

12.8-5. Suponga que se ha aplicado la técnica de programación separable a cierto problema (el "problema original") para convertirlo en el siguiente problema equivalente de programación lineal:

Maximizar $Z = 5x_{11} + 4x_{12} + 2x_{13} + 4x_{21} + x_{22}$,

sujeta a

$$3x_{11} + 3x_{12} + 3x_{13} + 2x_{21} + 2x_{22} \leq 25$$
$$2x_{11} + 2x_{12} + 2x_{13} - x_{21} - x_{22} \leq 10$$

y

$$0 \leq x_{11} \leq 2 \qquad 0 \leq x_{21} \leq 3$$
$$0 \leq x_{12} \leq 3 \qquad 0 \leq x_{22} \leq 1.$$
$$0 \leq x_{13}$$

¿Cuál era el modelo matemático del problema original? (Defina la función objetivo de manera algebraica o gráfica, pero exprese las restricciones en forma algebraica.)

12.8-6. *Pruebe* que, en cada uno de los siguientes casos, debe cumplirse la propiedad esencial de programación separable que se expuso

en la sección 12.8. (*Sugerencia:* Suponga que existe una solución óptima que viola esta propiedad y después contradiga esta suposición demostrando que existe una solución factible mejor.)

a) El caso especial donde las $g_i(\mathbf{x})$ son funciones lineales.

b) El caso general de programación separable donde todas las funciones son no lineales de la forma indicada. [*Sugerencia:* Considere las restricciones funcionales como restricciones sobre los recursos, donde $g_{ij}(x_j)$ representa la cantidad empleada del recurso *i* al realizar la actividad *j* en el nivel x_j; después, utilice la implicación del supuesto de convexidad sobre las pendientes de las funciones lineales por partes.]

12.8-7. La compañía MFG produce cierto subensamble en cada una de dos plantas. Estos subensambles se llevan a una tercera planta cercana donde se usan para fabricar determinado producto. Se acerca la temporada pico de la demanda de este producto, y para mantener las tasas de producción dentro de un margen deseado, es necesario usar temporalmente horas extra para hacer los subensambles. En la tabla que sigue se muestra el costo de cada uno de ellos en tiempo normal (TN) y en horas extra (HE) de ambas plantas, y el máximo número de subensambles que se puede producir en TN y en HE al día.

	Costo unitario		Capacidad	
	TN	**HE**	**TN**	**HE**
Planta 1	\$23	\$38	6 000	3 000
Planta 2	\$24	\$36	3 000	1 500

Sean x_1 y x_2 el número total de subensambles que se fabrica al día en las plantas 1 y 2, respectivamente. Suponga que el objetivo es maximizar $Z = x_1 + x_2$ sujeta a la restricción de que el costo diario total no puede exceder de 270 000 dólares. Observe que la formulación de programación matemática de este problema (con x_1 y x_2 como variables de decisión) tiene la misma forma que el caso principal de programación separable descrito en la sección 12.8, excepto que las funciones separables aparecen en una función de restricción y no en la función objetivo. No obstante, si se permite usar HE cuando aún la capacidad de TN en esa planta no está agotada, es posible usar el mismo enfoque para reformular el problema como uno de programación lineal.

a) Formule el modelo de programación lineal.

b) Explique por qué en este caso también se puede aplicar la lógica de programación separable para garantizar que una solución óptima del modelo formulado en el inciso *a*) nunca emplee HE, a menos que la capacidad de TN en esa planta esté agotada.

12.8-8. Considere el siguiente problema de programación no lineal

$$\text{Maximizar} \quad Z = 5x_1 + x_2,$$

sujeta a

$$2x_1^2 + x_2 \leq 13$$
$$x_1^2 + x_2 \leq 9$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

a) Demuestre que éste es un problema de programación convexa.

b) Utilice la técnica de programación separable de la sección 12.8 para formular un modelo aproximado de programación lineal. Use los enteros como puntos de salto de la función lineal por partes.

C ***c***) Utilice el método símplex para resolver el modelo que formuló en *b*); después, reexprese esta solución en términos de las variables originales del problema.

12.8-9. Considere el siguiente problema de programación convexa:

$$\text{Maximizar} \quad Z = 32x_1 - x_1^4 + 4x_2 - x_2^2,$$

sujeta a

$$x_1^2 + x_2^2 \leq 9$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

a) Aplique la técnica de programación separable que se presentó en la sección 12.8, con $x_1 = 0, 1, 2, 3$ y $x_2 = 0, 1, 2, 3$ como puntos de salto de las funciones lineales por partes para formular un modelo de programación lineal aproximado de este problema.

C ***b***) Utilice el método símplex para resolver el modelo que formuló en *a*). Después reescriba esta solución en términos de las variables *originales* del problema.

c) Aplique las condiciones KKT para determinar si la solución de las variables originales de *b*) es óptima para el problema original (no para el modelo aproximado).

12.8-10. Reconsidere el modelo de programación no lineal entera del problema 10.3-9.

a) Demuestre que la función objetivo no es cóncava.

b) Formule un modelo de programación *lineal* entera *binaria pura* para representar este problema, como sigue. Aplique la técnica de programación separable con los enteros como puntos de salto de las funciones lineales por partes, de forma que las variables auxiliares sean variables binarias. Después agregue restricciones de programación lineal sobre estas variables para reforzar la *restricción especial* de programación separable. (Observe que la *propiedad clave* de programación separable no se cumple porque la función objetivo no es cóncava.)

C ***c***) Utilice la computadora para resolver este problema según lo formuló en *b*). Después reescriba esta solución en términos de las variables *originales* del problema.

D,I **12.9-1.** Reconsidere el modelo de programación convexa linealmente restringida del problema 12.6-5. A partir de la solución de prueba inicial $(x_1, x_2) = (0, 0)$, realice una iteración del algoritmo de Frank-Wolfe para obtener exactamente la misma solución que encontró en el inciso *b*) del problema 12.6-5 y después ejecute una segunda iteración para verificar que se trata de una solución óptima (porque es una réplica exacta).

D,I **12.9-2.** Reconsidere el modelo de programación convexa linealmente restringida del problema 12.6-12. A partir de la solución de prueba inicial $(x_1, x_2) = (0, 0)$, realice una iteración del algoritmo de Frank-Wolfe para obtener exactamente la misma solución que encontró en el inciso *c*) del problema 12.6-12 y después lleve a cabo una segunda iteración para verificar que se trata de una solución óptima (porque es una réplica exacta). Explique por qué se obtendrían los mismos resultados en estas dos iteraciones con cualquier otra solución de prueba.

D,I **12.9-3.** Reconsidere el modelo de programación convexa linealmente restringida del problema 12.6-13. Comience con la solución de prueba inicial $(x_1, x_2, x_3) = (0, 0, 0)$ y luego realice dos iteraciones del algoritmo de Frank-Wolfe.

D,I **12.9-4.** Considere el ejemplo de programación cuadrática que se presentó en la sección 12.7. A partir de la solución de prueba inicial $(x_1, x_2) = (5, 5)$, aplique ocho iteraciones del algoritmo de Frank-Wolfe.

12.9-5. Reconsidere el modelo de programación cuadrática del problema 12.7-4.

D,I ***a***) Comience con la solución de prueba inicial $(x_1, x_2) = (0, 0)$ utilice el algoritmo de Frank-Wolfe (seis iteraciones) para resolver el problema (aproximado).

b) Grafique la forma en que se puede extrapolar la sucesión de soluciones de prueba que obtuvo en *a*) para obtener una aproximación más cercana a una solución óptima. ¿Qué estimación de esta solución obtuvo?

D,I **12.9-6.** Reconsidere el modelo de programación convexa linealmente restringida del problema 12.4-7. A partir de la solución de prueba inicial $(x_1, x_2) = (0, 0)$, utilice el algoritmo de Frank-Wolfe (cuatro iteraciones) para resolver este modelo (aproximado).

D,I **12.9-7.** Considere el siguiente problema de programación convexa linealmente restringida:

$$\text{Maximizar} \quad f(\mathbf{x}) = 3x_1x_2 + 40x_1 + 30x_2 - 4x_1^2 - x_1^4 - 3x_2^2 - x_2^4,$$

sujeta a

$$4x_1 + 3x_2 \leq 12$$
$$x_1 + 2x_2 \leq 4$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

A partir de la solución de prueba inicial $(x_1, x_2) = (0, 0)$, aplique dos iteraciones del algoritmo de Frank-Wolfe.

D,I **12.9-8.*** Considere el siguiente problema de programación convexa linealmente restringida:

$$\text{Maximizar} \quad f(\mathbf{x}) = 3x_1 + 4x_2 - x_1^3 - x_2^2,$$

sujeta a

$$x_1 + x_2 \leq 1$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

a) Comience con la solución de prueba inicial $(x_1, x_2) = (\frac{1}{4}, \frac{1}{4})$, para realizar tres iteraciones del algoritmo de Frank-Wolfe.

b) Utilice las condiciones KKT para verificar si la solución que obtuvo en *a*) en realidad es óptima.

12.9-9. Considere el siguiente problema de programación convexa linealmente restringida.

$$\text{Maximizar} \quad f(\mathbf{x}) = 4x_1 - x_1^4 + 2x_2 - x_2^2,$$

sujeta a

$$4x_1 + 2x_2 \leq 5$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

a) Comience con la solución de prueba inicial $(x_1, x_2) = (\frac{1}{2}, \frac{1}{2})$ para realizar cuatro iteraciones del algoritmo de Frank-Wolfe.

b) Demuestre en forma gráfica cómo se puede extrapolar la sucesión de soluciones de prueba que obtuvo en *a*) para obtener una aproximación más cercana a la solución óptima. ¿Cuál es la estimación que resulta de esta solución?

c) Utilice las condiciones KKT para verificar si la solución que obtuvo en *b*) es en realidad óptima. Si no lo es, utilice estas condiciones para derivar una solución óptima exacta.

12.9-10. Reconsidere el modelo de programación convexa linealmente restringido del problema 12.9-8.

a) Si se aplicara SUMT directamente a este problema, ¿cuál sería la función no restringida $P(\mathbf{x}; r)$ que debe minimizarse en cada iteración?

b) Establezca $r = 1$ y use $(\frac{1}{4}, \frac{1}{4})$ como solución de prueba inicial, para aplicar una iteración del procedimiento de búsqueda del gradiente (deténgase antes de despejar t^*) para maximizar la función $P(\mathbf{x}; r)$ que obtuvo en el inciso *a*).

D,C ***c***) A partir de la misma solución de prueba inicial del inciso *b*), use una computadora para aplicar la técnica SUMT a este problema con $r = 1, 10^{-2}, 10^{-4}$.

d) Compare el resultado del inciso *c*) de la solución óptima del problema 12.9-8. ¿Cuál es el porcentaje de error en x_1, en x_2 y en $f(\mathbf{x})$?

12.9-11. Reconsidere el modelo de programación convexa linealmente restringido del problema 12.9-9. Siga las instrucciones de los incisos *a*), *b*) y *c*) del problema 12.9-10 para este modelo, pero use $(x_1, x_2) = (\frac{1}{2}, \frac{1}{2})$ como solución de prueba inicial y $r = 1, 10^{-2}, 10^{-4}, 10^{-6}$.

12.9-12. Reconsidere el modelo del problema 12.3-3.

a) Si se aplicara SUMT directamente a este problema, ¿cuál sería la función no restringida $P(\mathbf{x}; r)$ que debe *minimizarse* en cada iteración?

b) Establezca $r = 100$ y use (x_1, x_2) (5, 5) como solución de prueba inicial para aplicar una iteración manual del procedimiento de búsqueda del gradiente (deténgase antes de despejar t^*) para minimizar la función $P(\mathbf{x}; r)$ que obtuvo en el inciso *a*).

D,C ***c***) A partir de la misma solución prueba inicial del inciso *b*), use el procedimiento automático del IOR Tutorial para aplicar la técnica SUMT a este problema con $r = 100, 1, 10^{-2}, 10^{-4}$. (*Sugerencia:* La rutina de computadora supone que el problema se ha convertido a la forma de maximización con restricciones funcionales de la forma $\leq$.)

12.9-13. Considere el ejemplo que se dio en la sección 12.9 con el objeto de aplicar la técnica SUMT.

a) Demuestre que $(x_1, x_2) = (1, 2)$ satisface las condiciones KKT.

b) Muestre la región factible en una gráfica, después localice los puntos $x_1x_2 = 2$ para demostrar que $(x_1, x_2) = (1, 2)$ con $f(1, 2) = 2$ es en realidad un *máximo global*.

12.9-14.* Considere el siguiente problema de programación convexa.

$$\text{Maximizar} \quad f(\mathbf{x}) = -2x_1 - (x_2 - 3)^2,$$

sujeta a

$$x_1 \geq 3 \quad \text{y} \quad x_2 \geq 3.$$

a) Si se aplicara la técnica SUMT a este problema, ¿cuál sería la función no restringida $P(\mathbf{x}; r)$ que debe maximizarse en cada iteración?
b) Derive la solución de maximización de $P(\mathbf{x}; r)$ analíticamente y después calcule esta solución con $r = 1, 10^{-2}, 10^{-4}, 10^{-6}$.
D,C ***c)*** A partir de la solución prueba inicial $(x_1, x_2) = (4, 4)$, utilice el procedimiento automático del IOR Tutorial para aplicar la técnica SUMT a este problema con $r = 1, 10^{-2}, 10^{-4}, 10^{-6}$.

D,C **12.9-15.** Considere el siguiente problema de programación convexa:

$$\text{Maximizar} \quad f(\mathbf{x}) = x_1x_2 - x_1 - x_1^2 - x_2 - x_2^2,$$

sujeta a

$$x_2 \geq 0.$$

A partir de la solución de prueba inicial $(x_1, x_2) = (1, 1)$, utilice el procedimiento automático del IOR Tutorial para aplicar SUMT a este problema, con $r = 1, 10^{-2}, 10^{-4}$.

D,C **12.9-16.** Reconsidere el modelo de programación cuadrática del problema 12.7-4. Comience con la solución de prueba inicial $(x_1, x_2) = (\frac{1}{2}, \frac{1}{2})$, luego, utilice el procedimiento automático del IOR Tutorial para aplicar SUMT al modelo con $r = 1, 10^{-2}, 10^{-4}, 10^{-6}$.

D,C **12.9-17.** Reconsidere la primera variación de programación cuadrática del problema de la Wyndor Glass de la sección 12.2 (vea la figura 12.6). A partir de la solución de prueba inicial $(x_1, x_2) = (2, 3)$, utilice el procedimiento automático del IOR Tutorial para aplicar SUMT a este problema con $r = 10^2, 1, 10^{-2}, 10^{-4}$.

12.9-18. Reconsidere el modelo de programación convexa con una restricción de igualdad del problema 12.6.11.
a) Si se aplicara la técnica SUMT a este problema, ¿cuál sería la función no restringida $P(\mathbf{x}; r)$ que debe minimizarse en cada iteración?
D,C ***b)*** A partir de la solución de prueba inicial $(x_1, x_2, x_3) = (\frac{3}{2}, \frac{3}{2}, 2)$, utilice el procedimiento automático del IOR Tutorial para aplicar la técnica SUMT a este problema con $r = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$.
C ***c)*** Utilice el Excel Solver estándar para resolver este problema.
C ***d)*** Utilice el Evolutionary Solver para resolver este problema.
C ***e)*** Utilice LINGO para resolver este problema.

12.10-1. Considere el siguiente problema de programación no convexa:

$$\text{Maximizar} \quad f(x) = 1\,000x - 400x^2 + 40x^3 - x^4,$$

sujeta a

$$x^2 + x \leq 500$$

y

$$x \geq 0.$$

a) Identifique los valores factibles de x. Obtenga las expresiones generales de las tres primeras derivadas de $f(x)$. Utilice esta información como ayuda para hacer un bosquejo de $f(x)$ sobre la región factible de x. Sin calcular sus valores, marque en la gráfica los puntos que corresponden a máximos y mínimos locales.
I ***b)*** Utilice el procedimiento de búsqueda de bisección con $\epsilon = 0.05$ para encontrar todos los máximos locales. Use el bosquejo del inciso *a*) para identificar las cotas iniciales apropiadas de cada una de estas búsquedas. ¿Cuál de los máximos locales es un máximo global?
c) A partir de $x = 3$ y $x = 15$ como las soluciones de prueba iniciales, utilice el método de Newton con $\epsilon = 0.001$ para encontrar cada uno de los máximos locales.
D,C ***d)*** Utilice el procedimiento automático del IOR Tutorial para aplicar SUMT a este problema, con $r = 10^3, 10^2, 10, 1$ para encontrar cada máximo local. Considere $x = 3$ y $x = 15$ como soluciones prueba iniciales para estas búsquedas. ¿Cuál de los máximos locales es un máximo global?
C ***e)*** Formule este problema en una hoja de cálculo, donde $f(x)$ representa ganancia, y después utilice la tabla de Solver para generar las soluciones con los siguientes puntos de inicio: $x = 0, 5, 10, 15, 20, 25$. Incluya el valor x y la ganancia como las celdas de salida de la tabla de Solver.
C ***f)*** Utilice el Evolutionary Solver para resolver este problema.
C ***g)*** Básese en la característica de optimización global de LINGO para resolver este problema.
C ***h)*** Emplee MPL y su optimizador global LGO para resolver este problema.

12.10-2. Considere el siguiente problema de programación no convexa:

$$\text{Maximizar} \quad f(\mathbf{x}) = 3x_1x_2 - 2x_1^2 - x_2^2,$$

sujeta a

$$x_1^2 + 2x_2^2 \leq 4$$
$$2x_1 - x_2 \leq 3$$
$$x_1x_2^2 + x_1^2x_2 = 2$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

a) Si se aplicara la SUMT al problema, ¿cuál sería la función no restringida $P(\mathbf{x}; r)$ que debe minimizarse en cada iteración?
D,C ***b)*** A partir de la solución de prueba inicial $(x_1, x_2) = (1, 1)$, utilice el procedimiento automático del IOR Tutorial para aplicar SUMT a este problema con $r = 1, 10^{-2}, 10^{-4}$.
C ***c)*** Utilice el Evolutionary Solver para resolver este problema.
C ***d)*** Use la característica de optimización global de LINGO para resolver este problema.
C ***e)*** Use MPL y su optimizador global para resolver este problema.

12.10-3. Considere el siguiente problema de programación no convexa.

$$\text{Minimizar} \quad f(\mathbf{x}) = \text{sen}\, 3x_1 + \cos 3x_2 + \text{sen}(x_1 + x_2),$$

sujeta a

$$x_1^2 - 10x_2 \geq -1$$
$$10x_1 + \quad x_2^2 \leq 100$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

a) Si se aplicara la SUMT al problema, ¿cuál sería la función no restringida $P(\mathbf{x}; r)$ que debe minimizarse en cada iteración?
b) Describa cómo podría aplicarse la SUMT para intentar obtener un mínimo global (no resuelva).
C ***c***) Utilice la característica de optimización global de LINGO para resolver este problema.
C ***d***) Básese en MPL y su optimizador global para resolver este problema.

C **12.10-4.** Considere el siguiente problema de programación no convexa:

$$\text{Maximizar ganancia} = x^5 - 13x^4 + 59x^3 - 107x^2 + 61x,$$

sujeta a

$$0 \leq x \leq 5.$$

a) Formule este problema en una hoja de cálculo y después utilice la tabla de Solver para generar las soluciones con los siguientes puntos de inicio: $x = 0, 1, 2, 3, 4$ y 5. Incluya el valor x y la ganancia como las celdas de salida en la tabla de Solver.
b) Utilice el Evolutionary Solver para resolver este problema.

C **12.10-5.** Considere el siguiente problema de programación no convexa:

$$\text{Maximizar ganancia} = 100x^6 - 1\,359x^5 + 6\,836x^4 - 15\,670x^3 + 15\,870x^2 - 5\,095x,$$

sujeta a

$$0 \leq x \leq 5.$$

a) Formule este problema en una hoja de cálculo y después utilice la tabla de Solver para generar las soluciones con los siguientes puntos de inicio: $x = 0, 1, 2, 3, 4$ y 5. Incluya el valor x y la ganancia como las celdas de salida en la tabla de Solver.
b) Utilice el Evolutionary Solver para resolver este problema.

C **12.10-6.** Debido al crecimiento de la población, el estado de Washington ha recibido un asiento adicional en la cámara de representantes, con lo que ahora cuenta con un total de 10 escaños. La Legislatura, que en la actualidad está controlada por los republicanos, necesita desarrollar un plan para redefinir los distritos en que se dividirá el estado. Existen 18 ciudades importantes en él que deben asignarse a uno de los 10 distritos del Congreso. En la tabla siguiente se proporcionan los números de demócratas y republicanos registrados en cada ciudad. Cada distrito debe contener entre 150 000 y 350 000 de estos votantes registrados. Use el Evolutionary Solver para asignar cada ciudad a uno de los 10 distritos del Congreso de forma que se maximice el número de distritos que tenga registrados más republicanos que demócratas. (*Sugerencia:* Utilice la función SUMIF.)

Ciudad	Demócratas (miles)	Republicanos (miles)
1	152	62
2	81	59
3	75	83
4	34	52
5	62	87
6	38	87
7	48	69
8	74	49
9	98	62
10	66	72
11	83	75
12	86	82
13	72	83
14	28	53
15	112	98
16	45	82
17	93	68
18	72	98

12.10-7. Reconsidere el problema de la Wyndor Glass Co., que se presentó en la sección 3.1.
C ***a***) Resuelva este problema mediante el uso del Excel Solver estándar.
C ***b***) A partir de una solución inicial de producir 0 lotes de puertas y 0 lotes de ventanas, resuelva este problema mediante el uso del Evolutionary Solver.
c) Comente acerca del desempeño de cada uno de los dos enfoques.

12.11-1. Considere el siguiente problema:

$$\text{Maximizar} \quad Z = 4x_1 - x_1^2 + 10x_2 - x_2^2,$$

sujeta a

$$x_1^2 + 4x_2^2 \leq 16$$

y

$$x_1 \geq 0, \quad x_2 \geq 0.$$

a) ¿Es éste un problema de programación convexa? Responda sí o no y justifique su respuesta.
b) ¿Se puede modificar el método símplex para usarlo en la solución de este problema? Responda sí o no y justifique su respuesta (pero no lo resuelva).
c) ¿Puede usarse el algoritmo de Frank-Wolfe para resolver este problema? Responda sí o no y después justifique su respuesta (pero no lo resuelva).
d) ¿Cuáles son las condiciones KKT de este problema? Utilice estas condiciones para determinar si $(x_1, x_2) = (1, 1)$ puede ser óptima.
e) Utilice la técnica de programación separable para formular un modelo de programación lineal *aproximado* para manejar este problema. Use los enteros factibles como puntos de salto de la función lineal por partes.
C ***f***) Utilice el método símplex para resolver el problema según se formuló en el inciso *e*).

g) Dada la función $P(\mathbf{x}; r)$, maximizar en cada iteración al aplicar SUMT al problema. (No resuelva.)

D,C ***h***) Use SUMT para resolver el problema que se formuló en *g*). Comience con la solución de prueba inicial $(x_1, x_2) = (2, 1)$ y utilice $r = 1, 10^{-2}, 10^{-4}, 10^{-6}$.

C ***i***) Formule este problema en una hoja de cálculo y después utilice el Excel Solver estándar para resolver este problema.

C ***j***) Use el Evolutionary Solver para resolver este problema.

C ***k***) Use LINGO para resolver este problema.

CASOS

Caso 12.1 Selección inteligente de acciones

Desde el día que tomó su primer curso de economía en la preparatoria, Lidia se preguntó sobre las prácticas financieras de sus padres, que trabajaban duro para ganar suficiente dinero para vivir una vida cómoda, pero nunca hicieron que su dinero trabajara para ellos. Sólo depositaban sus cheques ganados con esfuerzo en cuentas de ahorro que ganaban un interés nominal. (Por fortuna, siempre hubo suficiente dinero cuando llegó el momento de pagar la universidad.) Se prometió a sí misma que cuando llegara a la adultez, ella no seguiría las mismas prácticas conservadoras de sus padres.

Lidia cumplió su promesa. Cada mañana mientras se alista para el trabajo ve el reporte financiero de CNN. Participa en los juegos de inversión de internet, en donde encuentra carteras que maximizan su rendimiento y minimizan su riesgo. Lee *The Wall Street Journal* y el *Financial Times* con ansia de aprender.

Lidia también presta atención a las columnas de los consultores de inversiones de las revistas financieras y ha observado que, en promedio, los consejos que aparecen en ellas resultan muy buenos. Por lo tanto, decide seguir las sugerencias que presentan los últimos números de esas revistas. En su columna mensual, el editor Jonathan Taylor recomienda tres acciones que piensa que subirán muy por arriba del promedio del mercado. Además, la conocida gurú de fondos mutuos Donna Carter insiste en la compra de otras tres acciones que cree que tendrán el mejor desempeño del mercado durante el siguiente año.

BIGBELL (símbolo en el mercado de valores: BB), una de las compañías de telecomunicaciones más grandes del país, se vende con una tasa de ganancia por debajo del promedio del mercado. Grandes inversiones durante los últimos 8 meses han deprimido las ganancias de manera considerable. Sin embargo, debido a la nueva tecnología, se espera que la compañía eleve sus márgenes de utilidades en forma significativa. Taylor predice que las acciones subirán de su precio actual de 60 dólares cada una a 72 dólares en un año.

LOTSOFPLACE (LOP) es una de las más poderosas manufactureras del mundo. Hace poco la industria se consolidó, cuando una feroz guerra de precios durante los últimos años provocó que muchos competidores quebraran o vendieran sus compañías a LOTSOFPLACE y sus competidores. Debido a la menor competencia en el mercado, se espera que los ingresos y ganancias se eleven de manera considerable en el siguiente año. Taylor pronostica que, de su precio unitario actual de 127 dóalres, en un año las acciones de LOTSOFPLACE subirán 42%.

INTERNETLIFE (ILI) ha sobrevivido a las muchas altas y bajas de las compañías de internet. Con el furor por internet que se avecina, Taylor espera que la compañía duplique el precio de sus acciones de 4 a 8 dólares en un año.

HEALTHTOMORROW (HEAL) es una compañía biotecnológica líder que está a punto de obtener la aprobación de la FDA para varios medicamentos nuevos, lo que ayudará a que sus ganancias se eleven 20% en los próximos años. En particular, se supone que una nueva medicina para reducir significativamente el riesgo de ataques al corazón dará grandes ganancias. Además, debido al lanzamiento de varios medicamentos con sabor delicioso para niños, la compañía se ha forjado una excelente imagen en el medio. Sin duda, esta hazaña de relaciones públicas tendrá efectos positivos sobre sus ventas de medicamentos sin receta. Carter está convencida de que el precio unitario de las acciones se elevará de 50 a 75 dólares en un año.

QUICKY (QUI) es una cadena de restaurantes de comida rápida que se ha expandido por todo Estados Unidos. Carter ha seguido de cerca a esta compañía desde su entrada a la Bolsa hace cerca de 15 años cuando tenía unas docenas de restaurantes en la costa oeste. Desde entonces se ha expandido y ahora cubre todos los estados. Debido a su preocupación por la comida sana, capta un porcentaje de mercado creciente. Carter cree que las acciones continuarán con el buen desempeño arriba del promedio de mercado con un incremento de 46% en un año sobre el precio actual de 150 dólares.

AUTOMOBILE ALLIANCE (AUA) es el fabricante líder de automóviles en el área de Detroit que acaba de introducir dos modelos nuevos, los cuales muestran ventas iniciales muy fuertes. Por lo tanto, se predice que las acciones de la compañía subirán de 20 a 26 dólares en un año.

En internet, Lidia encontró datos acerca del riesgo involucrado en las acciones de estas compañías. Las varianzas históricas de los rendimientos de las seis acciones y sus covarianzas se muestran en las siguientes tablas.

a) Primero, Lidia quiere pasar por alto el riesgo de todas las inversiones. Dada esta estrategia, ¿cuál es su cartera de inversión óptima, es decir, qué fracción de su dinero debe invertir en cada una de las seis acciones? ¿Cuál es el riesgo total de su cartera?

b) Lidia decide que no quiere invertir más de 40% en una sola acción. Si aún pasa por alto el riesgo, ¿cuál es su nueva cartera de inversión óptima? ¿Cuál es el riesgo total de su nueva cartera?

c) Ahora Lidia quiere tomar en cuenta el riesgo de sus oportunidades de inversión. Para usarlo en los siguientes incisos, formule un modelo de programación cuadrática que maxi-

Compañía	**BB**	**LOP**	**ILI**	**HEAL**	**QUI**	**AUA**
Varianza	0.032	0.1	0.333	0.125	0.065	0.08

Covarianza	**LOP**	**ILI**	**HEAL**	**QUI**	**AUA**
BB	0.005	0.03	−0.031	−0.027	0.01
LOP		0.085	−0.07	−0.05	0.02
ILI			−0.11	−0.02	0.042
HEAL				0.05	−0.06
QUI					−0.02

mice su riesgo (medido por la varianza del beneficio que obtendrá de su cartera), mientras asegura que el beneficio que obtenga sea al menos tan grande como la elección de un valor mínimo aceptable.

***d*)** Lidia quiere asegurar que recibirá un rendimiento esperado de, al menos, 35%. Desea lograr su meta con un riesgo mínimo. ¿Qué cartera de inversión le permite hacerlo?

***e*)** ¿Cuál es el riesgo mínimo que Lidia puede lograr si desea un rendimiento esperado de 25%? ¿Y de al menos 40%?

***f*)** ¿Percibe usted problemas o desventajas en el enfoque de Lidia hacia su estrategia de inversión?

(*Nota:* Para su conveniencia, en el sitio en internet de este libro se proporciona un archivo de datos para este caso.)

RESUMEN DE CASOS ADICIONALES EN NUESTRO SITIO EN INTERNET (www.mhhe.com/hillier)

CASO 12.2 Inversiones internacionales

Un analista financiero posee algunos bonos alemanes que ofrecen incrementar su tasa de interés si se conservan hasta su madurez completa en tres años más. También se pueden negociar en cualquier momento para obtener el principal original más el interés generado. El gobierno alemán acaba de introducir un impuesto a la ganancia de capital sobre el ingreso por intereses mayor a cierto nivel, por lo que, en estas circunstancias, conservar los bonos hasta su madurez resulta menos atractivo. Por tanto, el analista tiene que determinar su estrategia óptima de inversión relativa a cuántos bonos debe vender en cada uno de los siguientes tres años bajo unos cuantos escenarios diferentes.

CASO 12.3 Promoción de un cereal para el desayuno, revisado

Este caso es una continuación del caso 3.4 que involucra una campaña de publicidad del nuevo cereal para desayuno de Super Gain Corporation. El análisis que se requiere para dicho caso conduce a la aplicación de programación lineal. Sin embargo, existen ciertos supuestos de programación lineal que son bastante cuestionables en esta situación. En particular, es claro que el supuesto de que la ganancia total por la introducción del nuevo producto es proporcional al número total de exposiciones debidas a la campaña de publicidad, es sólo una aproximación gruesa. Para perfeccionar el análisis, se deben formular, aplicar y comparar un modelo de programación no lineal general y un modelo de programación separable.

CAPÍTULO 13

Metaheurística

En algunos de los capítulos anteriores se han descrito algoritmos que pueden usarse para obtener una solución óptima para varios tipos de modelos de IO, entre los cuales se incluyen ciertas clases de modelos de programación lineal, programación entera y programación no lineal. Estos algoritmos han probado ser invaluables para abordar una amplia variedad de problemas prácticos. Sin embargo, este enfoque no siempre funciona. Algunos problemas (y los correspondientes modelos de IO) pueden ser tan complicados que no es posible resolverlos para encontrar una solución óptima. En tales situaciones, aún es importante encontrar una buena solución factible que al menos esté razonablemente cerca de ser óptima. Por lo general, para buscar esa solución se utilizan métodos heurísticos.

Un **método heurístico** es un procedimiento que trata de descubrir una solución factible muy buena, pero no necesariamente una solución óptima, para el problema específico bajo consideración. No puede darse una garantía acerca de la calidad de la solución que se obtiene, pero un método heurístico bien diseñado puede proporcionar una solución que al menos está cerca de ser óptima (o concluir que no existen tales soluciones). El procedimiento también debe ser suficientemente eficiente como para manejar problemas muy grandes. Con frecuencia, el procedimiento es un *algoritmo iterativo* novedoso, donde cada iteración implica la realización de una búsqueda de una nueva solución que puede ser mejor que la solución que se encontró con anterioridad. Cuando el algoritmo termina después de un tiempo razonable, la solución que proporciona es la mejor que se pudo encontrar en cualquier iteración.

Con frecuencia los métodos heurísticos se basan en ideas bastante simples, de sentido común, acerca de la forma en que se debe buscar una buena solución. Estas ideas deben ajustarse al problema específico de interés. En consecuencia, los métodos heurísticos tienden a ser *ad hoc* por naturaleza. Esto es, por lo general cada método se diseña para abordar un tipo específico de problema en vez de una variedad de aplicaciones.

Por muchos años esta situación significó que, para desarrollar un método heurístico, un equipo de IO necesitaría comenzar desde cero para así ajustarse al problema bajo consideración, siempre que no exista un algoritmo disponible para encontrar una solución óptima. Este panorama ha cambiado en años relativamente recientes con el desarrollo de metaheurísticas poderosas. Una **metaheurística** es un método de solución general que proporciona tanto una estructura general como criterios estratégicos para desarrollar un método heurístico específico que se ajuste a un tipo particular de problema. La metaheurística se ha convertido en una de las técnicas más importantes del paquete de herramientas que utilizan los profesionales de la IO.

En este capítulo se proporciona una introducción elemental a la metaheurística. En la primera sección se describirá la naturaleza general de la metaheurística. Después, en las siguientes tres secciones, se presentarán e ilustrarán las tres metaheurísticas que se utilizan con mayor frecuencia.

13.1 NATURALEZA DE LA METAHEURÍSTICA

Para ilustrar la naturaleza de la metaheurística se iniciará con un ejemplo de un problema de programación lineal pequeño pero moderadamente difícil.

Ejemplo: Un problema de programación no-lineal con óptimos locales múltiples

Considere el siguiente problema:

$$\text{Maximizar} \quad f(x) = 12x^5 - 975x^4 + 28\,000x^3 - 345\,000x^2 + 1\,800\,000x,$$

sujeto a

$$0 \leq x \leq 31.$$

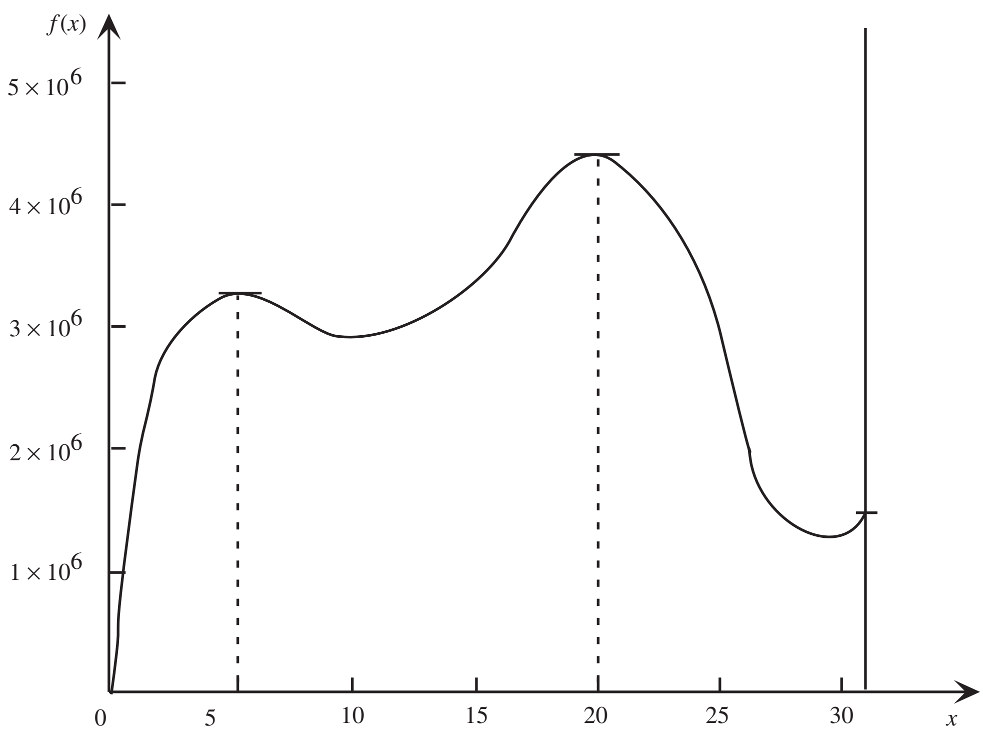
En la figura 13.1 se grafica la función objetivo $f(x)$ sobre los valores factibles de la variable única x. Esta gráfica revela que el problema tiene tres óptimos locales, uno en $x = 5$, otro en $x = 20$ y el tercero en $x = 31$, donde el óptimo global se encuentra en $x = 20$.

La función objetivo $f(x)$ es tan complicada que sería difícil determinar dónde se encuentra el óptimo global sin el beneficio de observar la gráfica de la figura 13.1. Se podría usar el cálculo, pero esto requeriría resolver una ecuación polinomial de cuarto grado (después de igualar a cero la primera derivada) para determinar dónde se encuentran los puntos críticos. Incluso sería difícil determinar que $f(x)$ tiene óptimos locales múltiples en vez de un solo óptimo global.

Este problema es un ejemplo de un problema de *programación no convexa*, un tipo especial de programación no lineal que por lo general tiene óptimos locales múltiples. En la sección 12.10 se explica la programación convexa e incluso se introduce un paquete de software (Solver evolutivo) que utiliza el tipo de metaheurística que se describe en la sección 13.4.

Para problemas de programación no lineal que parecen ser muy difíciles, como éste, un método heurístico común es ejecutar un **procedimiento de mejora local**. Tal procedimiento comienza con una solución de prueba inicial y después, en cada iteración, busca en la vecindad de la solución de prueba para tratar de encontrar una mejor solución que la actual. Este proceso continúa hasta que no se pueda encontrar una solución mejorada en la vecindad de la solución de prueba actual. De esta manera, este tipo de procedimiento puede verse como un procedimiento de *escalada de montaña* que se mantiene en ascenso en la gráfica de la función objetivo (si el objetivo es una maximización) hasta que, en esencia, alcanza la cumbre de la montaña. Un procedimiento de mejora local bien diseñado, por lo general convergirá hacia un óptimo *local* (una cumbre de montaña), pero se detendrá aun si este óptimo local no es un óptimo *global* (la cumbre más alta de la montaña).

FIGURA 13.1
Gráfica del valor de la función objetivo en un intervalo factible, $0 \leq x \leq 31$, del ejemplo de programación no lineal. Los óptimos locales están en $x = 5$, $x = 20$ y $x = 31$, pero sólo $x = 20$ es un óptimo global.



Por ejemplo, el *procedimiento de búsqueda de gradiente* descrito en la sección 12.5 es un procedimiento de mejora local. Si éste comienza, por ejemplo, con $x = 0$ como la solución de prueba inicial en la figura 13.1, ascendería por la pendiente al probar en forma exitosa valores más grandes de x hasta que, en esencia, llegara a la cumbre de la montaña en $x = 5$, punto en el que se detendría. En la figura 13.2 se muestra una secuencia típica de valores de $f(x)$ que se obtendría mediante este procedimiento de mejora local cuando se inicia desde la parte baja de la gráfica.

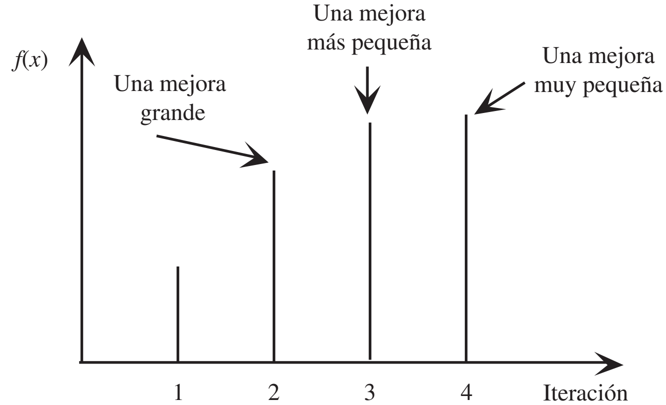
Como el ejemplo de programación no lineal que se muestra en la figura 13.1 involucra sólo una variable, el método de bisección que se describió en la sección 12.4 también se podría aplicar a este problema particular. Éste es otro ejemplo de un procedimiento de mejora local, puesto que cada iteración comienza en la solución de prueba actual a buscar una mejor solución en su vecindad (que se define por las cotas inferior y superior actuales sobre el valor de la variable). Por ejemplo, si la búsqueda comenzara con una cota inferior de $x = 0$ y una cota superior de $x = 6$ en la figura 13.1, la secuencia de soluciones de prueba que se obtendría por el método de bisección sería $x = 3$, $x = 4.5$, $x = 5.25$, $x = 4.875$, hasta convergir a $x = 5$. Los valores correspondientes de la función objetivo de estas cuatro soluciones de prueba son 2.975 millones, 3.286 millones, 3.300 millones y 3.302 millones, respectivamente. Por consiguiente, la segunda iteración proporciona una mejora más o menos grande sobre la primera (311 000), la tercera iteración da una mejora mucho más pequeña (14 000), mientras que la cuarta sólo ofrece una mejora marginal (2 000). Como se muestra en la figura 13.2, este patrón es bastante típico de los procedimientos de mejora local (aunque con alguna variación en la tasa de convergencia al máximo local).

Igual que en el procedimiento de búsqueda de gradiente, esta búsqueda con el método de bisección se quedaría atrapada en el óptimo local de $x = 5$, por lo que nunca encontraría el óptimo global en $x = 20$. Como otros procedimientos de mejora local, tanto el de búsqueda de gradiente como el método de bisección fueron diseñados sólo para obtener una mejoría sobre la solución de prueba actual dentro de su vecindad local. Una vez que éstos ascienden hasta el pico de la montaña, se deben detener porque ya no pueden ascender más dentro de la vecindad local de la solución de prueba en la cumbre. Este inconveniente ilustra la desventaja de cualquier procedimiento de mejora local.

Desventaja de un procedimiento de mejora local: Cuando se aplica un procedimiento de mejora local bien diseñado a un problema de optimización con múltiples óptimos locales, el procedimiento convergirá hacia un óptimo local y se detendrá. El óptimo local que encuentre dependerá del punto en el que el procedimiento inicia su búsqueda. Por lo tanto, el procedimiento encontrará el óptimo global sólo si inicia la búsqueda en la vecindad de dicho óptimo global.

Para tratar de evitar esta desventaja se puede reiniciar el procedimiento de mejora local cierto número de veces a partir de soluciones de prueba aleatorias. Con frecuencia, cuando se reinicia desde una parte nueva de la región factible se llega a un nuevo óptimo local. Si se repite esta rutina cierta cantidad de veces se incrementa la posibilidad de que el mejor óptimo local que se obtuvo en realidad sea el óptimo global. Este enfoque funciona bien con problemas pequeños, como el

■ **FIGURA 13.2**
Secuencia típica de valores de la función objetivo de las soluciones obtenidas por un procedimiento de mejora local, cuando éste converge hacia un óptimo local durante su aplicación a un problema de maximización.



problema de programación no lineal de una sola variable de la figura 13.1. Sin embargo, es mucho menos exitoso en problemas grandes con muchas variables y una región factible complicada. Cuando la región factible tiene muchos "recovecos" y sólo al reiniciar un procedimiento de mejora local en uno de ellos conducirá al óptimo global, la aplicación del procedimiento a partir de soluciones iniciales de prueba aleatorias se convierte en una forma muy fortuita de llegar al óptimo global.

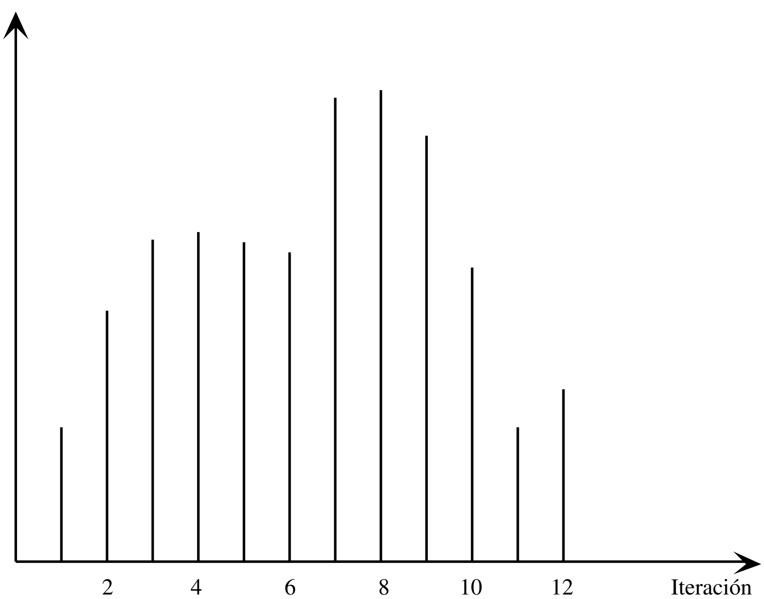
Lo que se necesita en lugar de esta mecánica es un enfoque más estructurado que utilice la información recopilada para dirigir la búsqueda hacia el óptimo global. Éste es el papel que tiene la metaheurística.

> **Naturaleza de la metaheurística:** La metaheurística es un tipo general de método de solución que organiza la interacción entre los procedimientos de mejora local y las estrategias de más alto nivel para crear un proceso que sea capaz de escapar de un óptimo local y realizar una búsqueda vigorosa de una región factible.

En este contexto, una característica clave de la metaheurística es la capacidad de escapar de un óptimo local. Después de encontrar (o casi encontrar) un óptimo local, las diferentes metaheurísticas ejecutan este escape de modos diferentes. Sin embargo, una característica común es que las soluciones de prueba que de inmediato siguen un óptimo local pueden ser menores que éste. En consecuencia, cuando se aplica una metaheurística a un problema de maximización (como en el ejemplo que se muestra en la figura 13.1), los valores de la función objetivo de la secuencia de soluciones de prueba que se obtengan por lo general seguirán un patrón similar al que se muestra en la figura 13.3. Como con la figura 13.2, el proceso comienza con la utilización de un procedimiento de búsqueda local para ascender a la cumbre actual (iteración 4). Sin embargo, en lugar de detenerse aquí, la metaheurística podría conducir la búsqueda un poco abajo hacia el otro lado de la cumbre hasta poder comenzar a ascender hasta la cumbre más alta (iteración 8). Para verificar que éste parece ser el óptimo global, una metaheurística continúa con la exploración antes de detenerse (iteración 12).

En la figura 13.3 se ilustra tanto la ventaja como la desventaja de una metaheurística bien diseñada. La ventaja es que tiende a moverse relativamente rápido hacia soluciones muy buenas, por lo que proporciona una forma muy eficiente de abordar problemas grandes y complicados. La desventaja es que no existe garantía de que la mejor solución que se encuentre sea una solución óptima o incluso que esté cerca de serlo. Por lo tanto, siempre que sea posible resolver un problema mediante un algoritmo que pueda garantizar optimalidad, debe usarse éste en lugar de la metaheurística. El papel de las metaheurísticas es abordar problemas que son muy grandes y complicados como para resolverlos por medio de algoritmos exactos. Todos los ejemplos de este

■ **FIGURA 13.3**
Secuencia típica de valores de la función objetivo de las soluciones que se obtuvieron por una metaheurística, cuando ésta converge primero hacia un óptimo local (iteración 4) y después escapa para hacerlo hacia el óptimo global (iteración 8) de un problema de maximización, antes de concluir su búsqueda (iteración 12).



capítulo son demasiado pequeños como para requerir el uso de las metaheurísticas, ya que sólo intentan ilustrar de una manera directa la forma en que las metaheurísticas pueden enfrentar problemas mucho más complicados.

En la sección 13.3 se ilustrará la aplicación de una metaheurística particular al ejemplo de programación no lineal que se muestra en la figura 13.1. Después, en la sección 13.4 se aplicará otra metaheurística a la versión para programación entera del mismo ejemplo.

Aunque algunas veces se aplican metaheurísticas a problemas difíciles de programación no lineal y entera, un área de aplicación más común es la de los problemas de *optimización combinatoria*. El siguiente ejemplo es de este tipo.

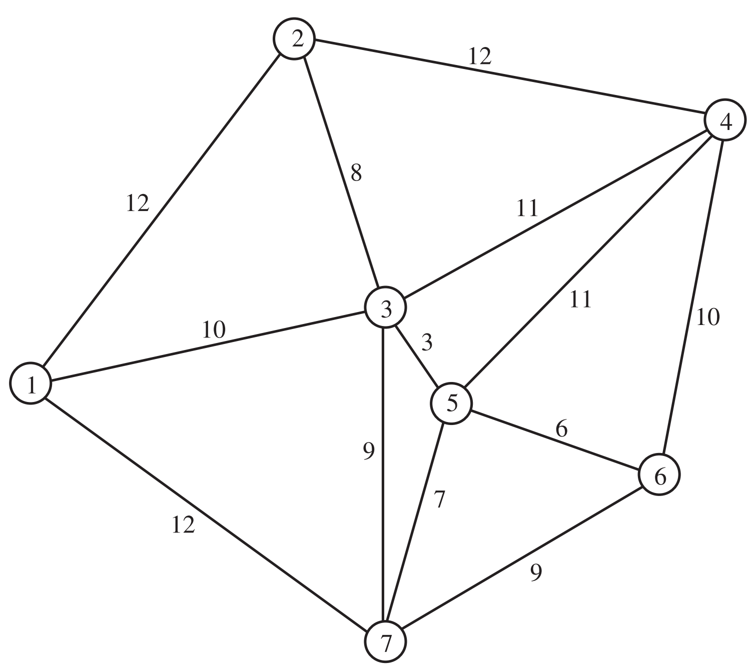
Ejemplo: Problema del agente viajero

Quizá el problema clásico de optimización combinatoria más famoso es conocido como el *problema del agente viajero*. Recibió este nombre pintoresco porque puede describirse en términos de un agente de ventas que debe visitar cierta cantidad de ciudades en un solo viaje. Si comienza desde su ciudad de residencia, el agente determinará cuál ruta debe seguir para visitar cada ciudad exactamente una vez antes de regresar a su casa de manera que se minimice la longitud total del viaje.

En la figura 13.4 se muestra un pequeño problema del agente viajero con siete ciudades. La ciudad 1 es el lugar de residencia del agente. Por lo tanto, si comienza desde su ciudad, el agente debe elegir una ruta para visitar cada una de las otras ciudades exactamente una vez antes de regresar a su punto de partida. El número colocado junto a la ligadura entre cada par de ciudades representa la distancia (o el costo o el tiempo) entre estas ciudades. Se supone que la distancia es la misma en cualquier dirección. (Esto se conoce como el problema *simétrico* del agente viajero.) Aunque por lo general existe una ligadura directa entre cada par de ciudades, en este caso el ejemplo se simplifica al suponer que las únicas ligaduras directas son las que se muestran en la figura. El objetivo es determinar cuál ruta minimizará la distancia total que el agente viajero debe viajar.

Ha existido una serie de aplicaciones del problema del agente viajero a problemas que no tienen nada que ver con un agente de ventas. Por ejemplo, cuando un camión sale de un centro de distribución para entregar bienes a cierta cantidad de ubicaciones, el problema de determinar la ruta más corta es un problema del agente viajero. Otro ejemplo involucra la fabricación de tableros de circuitos impresos, semiconductores cableados y otros componentes. Cuando es necesario perforar muchos orificios en un tablero de circuitos impresos, el problema de encontrar la secuencia de perforaciones más eficiente constituye un problema del agente viajero.

FIGURA 13.4
Ejemplo del problema del agente viajero que se usará con propósitos ilustrativos a lo largo de este capítulo.



El problema del agente viajero se complica con rapidez a medida que aumenta el número de ciudades. En el caso de un problema con n ciudades y una ligadura entre cada par de ellas, el número de rutas factibles que debe considerarse es $(n-1)!/2$ puesto que hay $(n-1)$ posibilidades para la primera ciudad después de la ciudad de residencia del agente, $(n-2)$ posibilidades para la siguiente ciudad y así sucesivamente. El denominador de 2 surge porque cada ruta tiene una vía inversa equivalente con la misma distancia. Así, mientras un problema del agente viajero con 10 ciudades tiene menos de 200 000 soluciones factibles que deben ser consideradas, un problema con 20 ciudades tiene alrededor de 10^{16} soluciones factibles, mientras que un problema con 50 ciudades tiene alrededor de 10^{62}.

Resulta sorprendente, pero algunos algoritmos poderosos basados en el enfoque de ramificación y acotamiento que se introdujo en la sección 11.8 han tenido éxito en resolver ciertos problemas del agente viajero muy grandes con muchos cientos (o incluso miles) de ciudades. Sin embargo, debido a la enorme dificultad para resolver grandes problemas de este tipo, los métodos heurísticos guiados por metaheurísticas son una forma popular de enfrentar tales problemas.

Por lo general, estos métodos heurísticos implican la generación de una secuencia de soluciones de prueba factibles, donde cada nueva solución de prueba se obtiene al hacer cierto tipo de pequeños ajustes en la solución de prueba actual. Se han sugerido algunos métodos para ajustar la solución de prueba actual. Debido a la facilidad de su implantación, un método popular usa el siguiente tipo de ajuste.

> Un **subviaje inverso** ajusta la secuencia de ciudades visitadas en la solución de prueba actual mediante la selección de una subsecuencia de las ciudades y simplemente invertir el orden en el cual se visita esa subsecuencia de ciudades. (La subsecuencia invertida puede consistir cuando menos en dos ciudades.)

Para ilustrar un subviaje inverso, suponga que la solución de prueba inicial del ejemplo de la figura 13.4 es visitar las ciudades en orden numérico:

1-2-3-4-5-6-7-1 Distancia = 69

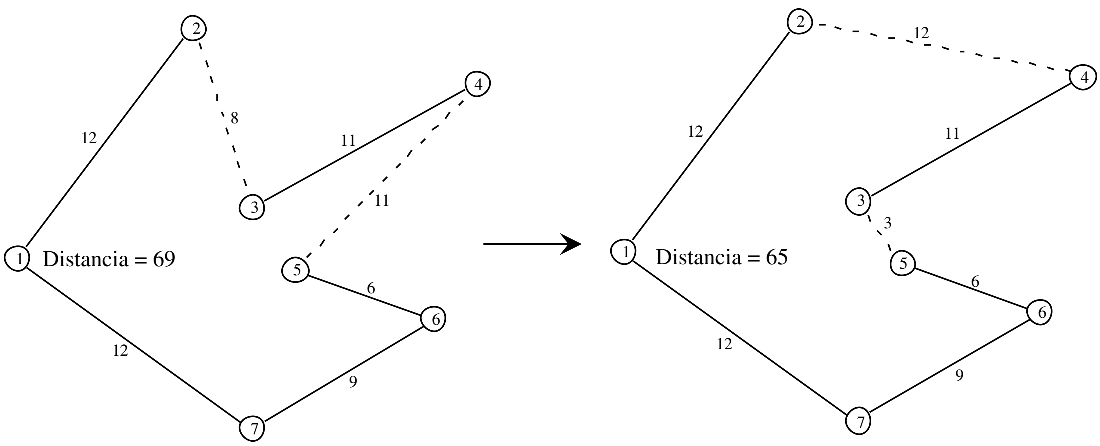
Por ejemplo, si se selecciona la secuencia 3-4 y se invierte, se obtiene la siguiente nueva solución de prueba:

1-2-4-3-5-6-7-1 Distancia = 65

En consecuencia, este subviaje inverso en particular ha tenido éxito en reducir la distancia del viaje completo de 69 a 65.

En la figura 13.5 se muestra este subviaje inverso, el cual conduce de la solución de prueba inicial de la izquierda a la nueva solución de prueba de la derecha. Las líneas punteadas indican las

FIGURA 13.5
Inversión de un subviaje que reemplaza al viaje de la izquierda (la solución de prueba inicial) por el viaje de la derecha (la nueva solución de prueba) al invertir el orden en que se visitan las ciudades 3 y 4. Esta inversión del subviaje resulta en la sustitución de las líneas punteadas de la izquierda por las líneas punteadas de la derecha como las ligaduras que se invierten en el nuevo viaje.



ligaduras que se borran del viaje (a la izquierda) o se agregan al viaje (a la derecha) por el subviaje inverso. Observe que la nueva solución de prueba elimina exactamente dos ligaduras del viaje previo y las reemplaza por exactamente dos nuevas ligaduras para formar el nuevo viaje. Ésta es una característica de cualquier subviaje inverso (incluso aquéllos donde la subsecuencia de ciudades invertida consiste en más de dos ciudades). Por consiguiente, un subviaje inverso particular es posible sólo si las dos nuevas ligaduras correspondientes en realidad existen.

Este éxito en la obtención de un viaje mejorado por la simple inversión de un subviaje sugiere el siguiente método heurístico para buscar una buena solución factible para cualquier problema del agente viajero.

Algoritmo del subviaje inverso

Paso inicial. Inicie con cualquier viaje factible como la solución de prueba inicial.

Iteración. Para la solución de prueba actual, considere todas las formas posibles de realizar un subviaje inverso (sólo excluya el inverso del viaje completo). Seleccione aquel que proporcione la mayor disminución en la distancia viajada para que sea la nueva solución de prueba. (Los empates se rompen de manera arbitraria.)

Regla de detención. El proceso se detiene cuando ninguna inversión de un subviaje mejora la solución de prueba actual, la cual se acepta como la solución final.

Ahora se aplicará este algoritmo al ejemplo. Se inicia con 1-2-3-4-5-6-7-1 como la solución de prueba inicial. Existen cuatro posibles subviajes que mejorarían esta solución, como a continuación se presenta en el segundo, tercero, cuarto y quinto renglones.

	1-2-3-4-5-6-7-1	Distancia = 69
inverso 2-3:	1-3-2-4-5-6-7-1	Distancia = 68
inverso 3-4:	1-2-4-3-5-6-7-1	Distancia = 65
inverso 4-5:	1-2-3-5-4-6-7-1	Distancia = 65
inverso 5-6:	1-2-3-4-6-5-7-1	Distancia = 66

Las dos soluciones con distancia = 65 empatan pues proporcionan la mayor disminución en la distancia viajada; luego, suponga que se elige de manera arbitraria la primera de éstas, 1-2-4-3-5-6-7-1 (como se muestra en el lado derecho de la figura 13.5), como la siguiente solución de prueba, con lo cual se completa la primera iteración.

La segunda iteración comienza con el viaje que se muestra en el lado derecho de la figura 13.5 como la solución de prueba actual. En el caso de esta solución, existe sólo un subviaje que proporciona una mejora, como se menciona a continuación en el segundo renglón:

	1-2-4-3-5-6-7-1	Distancia = 65
inverso 3-5-6:	1-2-4-6-5-3-7-1	Distancia = 64

En la figura 13.6 se muestra este subviaje inverso, donde la subsecuencia de ciudades 3-5-6 de la izquierda ahora se modifica en orden inverso (6-5-3) a la derecha. Por lo tanto, en el viaje a la derecha ahora se presenta la ligadura 4-6 en lugar de 4-3, así como la ligadura 3-7 en lugar de 6-7, para usar el orden inverso 6-5-3 entre las ciudades 4 y 7. Con lo anterior se completa la segunda iteración.

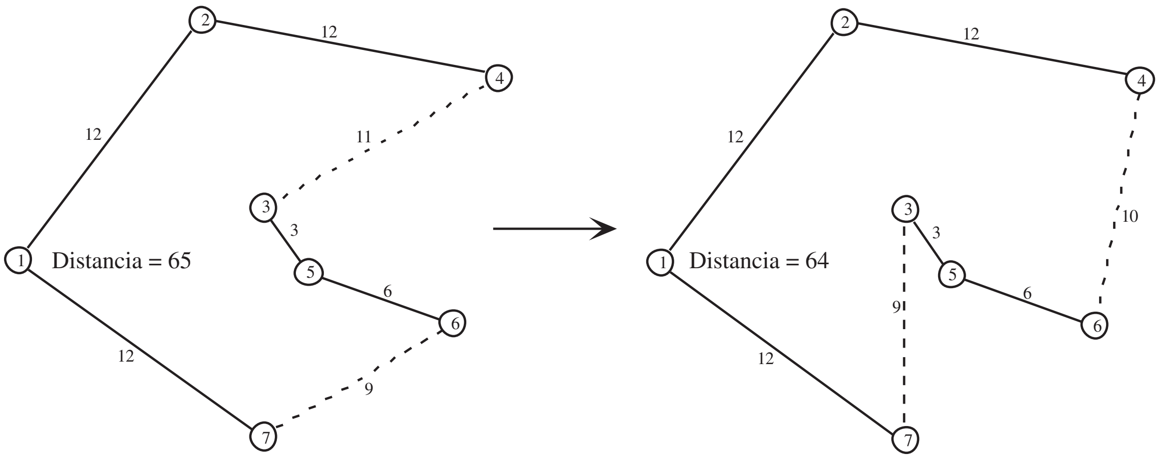
A continuación se tratará de encontrar el inverso que mejorará esta nueva solución de prueba. Sin embargo, no existe ninguna, por lo que se detiene el algoritmo del subviaje inverso con esta solución de prueba como la solución final.

¿La solución óptima es 1-2-4-6-5-3-7-1? Desafortunadamente no, pues la solución óptima resulta ser

1-2-4-6-7-5-3-1 Distancia = 63

(o 1-3-5-7-6-4-2-1 al invertir la dirección de este viaje completo)

Sin embargo, no se puede llegar a esta solución al realizar la inversión de un subviaje que mejore 1-2-4-6-5-3-7-1.



■ **FIGURA 13.6**
Inversión del subviaje 3-5-6 que conduce desde la solución de prueba de la izquierda hasta una solución de prueba mejorada de la derecha.

El algoritmo del subviaje inverso es otro ejemplo de un *procedimiento de mejora local*. Encuentra una mejoría sobre la solución de prueba actual en cada iteración. Cuando ya no puede encontrar una mejor solución se detiene porque la solución de prueba actual es un óptimo local. En este caso, en realidad 1-2-4-6-5-3-7-1 es un *óptimo local* porque no hay una mejor solución dentro de su vecindad local que pueda ser alcanzada mediante la inversión de un subviaje.

Lo que podría dar una mejor oportunidad de encontrar un óptimo local es utilizar una metaheurística que permita el proceso de escapar de un óptimo local. En las tres secciones siguientes se verá cómo tres metaheurísticas distintas logran este resultado con el mismo ejemplo.

13.2 BÚSQUEDA TABÚ

La búsqueda tabú es una metaheurística muy usada que utiliza algunas ideas de sentido común para permitir que el proceso de búsqueda escape de un óptimo local. Después de introducir los conceptos básicos, se abordará un ejemplo simple y después se regresará al ejemplo del agente viajero.

Conceptos básicos

Cualquier aplicación de la búsqueda tabú incluye como una subrutina algún *procedimiento de búsqueda local* que parezca apropiado para el problema bajo consideración. (Un **procedimiento de búsqueda local** opera como un procedimiento de mejora local excepto que no requiere que cada nueva solución de prueba sea mejor que la solución de prueba anterior.) El proceso comienza con este método como un procedimiento de *mejora* local de la manera usual (es decir, al aceptar sólo una solución mejorada en cada iteración) para encontrar un óptimo local. Una estrategia de la búsqueda tabú es que continúa la búsqueda pero permite *movimientos sin mejora* hacia las mejores soluciones en la vecindad del óptimo local. Una vez que se alcanza un punto en el que se pueden encontrar mejores soluciones en la vecindad de la solución de prueba, se aplica de nuevo el procedimiento de mejora local para encontrar un nuevo óptimo local.

Debido a la analogía con la escalada de una montaña, algunas veces este proceso se conoce como **enfoque del ascenso más empinado/descenso más suave** porque cada iteración selecciona el movimiento disponible que sube más en la pendiente o, cuando no hay disponible algún movimiento hacia arriba, selecciona el movimiento que baja menos en la pendiente. Si todo sale bien, el proceso seguirá un patrón como el que se muestra en la figura 13.3, donde se deja atrás un óptimo local con la intención de escalar hacia el óptimo global.

El peligro de este enfoque es que después de dejar un óptimo local, el proceso se puede ciclar y regresar al mismo óptimo local. Para evitar este círculo vicioso, la búsqueda tabú prohíbe en forma

Recuadro de aplicación

Sears, Roebuck and Company (comúnmente conocida con el nombre de **Sears**), fundada en 1886, creció hasta convertirse en la más grande tienda minorista multilínea en Estados Unidos a mediados del siglo XX. En la actualidad es uno de los minoristas más grandes en el mundo en la venta de mercancías y servicios. Asimismo, cuenta con el servicio más grande de entrega a domicilio de muebles y electrodomésticos de Estados Unidos, donde realiza más de 4 millones de entregas al año. En este país, esta tienda controla una flota de más de 1 000 vehículos para entrega de mercancía, en la cual se incluyen vehículos contratados y propios. Además cuenta, en Estados Unidos, con una flota de alrededor de 12 500 vehículos de servicio y los técnicos correspondientes, quienes realizan alrededor de 15 millones de llamadas de servicio a domicilio al año para reparar e instalar electrodomésticos, así como para brindar servicios de mejoras en el hogar.

El costo asociado con la operación de este enorme sistema de entrega a domicilio y servicio a los hogares se encuentra en el orden de *miles de millones de dólares al año*. Debido a los miles de vehículos que se utilizan para atender a las solicitudes de decenas de miles de llamadas de clientes *diariamente*, la eficiencia de esta operación tiene un efecto sorprendente en la rentabilidad de la compañía.

Para poder atender tantas llamadas de servicio por parte de los clientes con tantos vehículos, *cada* día es necesario tomar un enorme número de decisiones. ¿Qué paradas se le debe asignar a cada ruta? ¿Cuál debe ser el orden de dichas paradas (lo cual tiene un efecto significativo en la distancia y tiempo totales de la ruta) de cada vehículo? ¿Cómo pueden tomarse todas estas decisiones de tal manera que se minimicen los costos totales de operación y a la vez se ofrezca un servicio satisfactorio a los clientes?

Fue evidente que era necesario que la investigación de operaciones considerara este problema. La formulación natural del problema debía asumir la forma de un *problema de enrutamiento de vehículos con ventanas de tiempo* (VRPTW, vehicle-routing problem with time windows), para el cual se han desarrollado algoritmos exactos y heurísticos. Desafortunadamente, el problema de Sears es tan grande que representa un problema de optimización combinatoria muy difícil que está más allá del alcance de los algoritmos estándar de VRPTW. Por esta razón se desarrolló un nuevo algoritmo basado en el uso de la *búsqueda tabú* con el fin de tomar decisiones acerca de qué ruta vehicular atenderá qué paradas y cuál será la secuencia de paradas dentro de una cierta ruta.

El nuevo sistema de enrutamiento y programación de vehículos, el cual se basó en gran medida en la búsqueda tabú, representó para Sears un *ahorro por evento de más de* **9 millones de dólares** y *ahorros anuales* de *más de* **42 millones de dólares**. Asimismo, brindó un gran número de beneficios intangibles en los que se incluye un *mejor servicio* al cliente.

Fuente: D. Weigel y B. Cao: "Applying GIS and OR Techniques to Solve Sears Technician-Dispatching and Home-Delivery Problems", en *Interfaces*, **29**(1): 112-130, enero-febrero de 1999. (En el sitio en internet de este libro —www.mhhe.com/hillier— se proporciona una liga hacia este artículo.)

temporal los movimientos que pudieran regresar el proceso a una solución visitada recientemente. Una **lista tabú** registra estos movimientos prohibidos, los cuales se conocen como *movimientos tabú*. (La única excepción a la prohibición de un movimiento de este tipo se presenta cuando un movimiento tabú en realidad es mejor que la mejor solución factible que se haya encontrado hasta ese momento.)

Este uso de *memoria* para dirigir la búsqueda al utilizar listas que registran cierta parte de la historia reciente es una característica distintiva de la búsqueda tabú. Esta condición tiene raíces en el campo de la inteligencia artificial.

La búsqueda tabú también puede incorporar algunos conceptos avanzados. Uno de ellos es la *intensificación*, que implica la exploración de una parte de la región factible con más intensidad de la usual después de que se haya identificado como una parte particularmente promisoria para encontrar en ella muy buenas soluciones. Otro concepto es la *diversificación*, que implica forzar la búsqueda por medio de la entrada en áreas de la región factible que no se han explorado con anterioridad. (Para implantar ambos conceptos se utiliza la memoria de largo plazo.) Sin embargo, el enfoque aquí se centrará sobre la forma básica de la búsqueda tabú que se resume más adelante sin ahondar en estos conceptos adicionales.

Esquema de un algoritmo de búsqueda tabú básico

Paso inicial. Comience con una solución de prueba inicial factible.

Iteración. Utilice un procedimiento de búsqueda local apropiado para definir los movimientos factibles en la vecindad local de la solución de prueba actual. No considere la realización de ningún movimiento incluido en la lista tabú actual a menos que ese movimiento genere una mejor solución que la mejor solución de prueba que se haya encontrado hasta ahora. Determine cuál de los movimientos restantes proporciona la mejor solución. Adopte esta solución como la próxima solución

de prueba, sin importar si ésta es mejor o peor que la solución de prueba actual. Actualice la lista tabú para evitar el regreso a la última solución de prueba actual. Si la lista tabú ya está llena, elimine el elemento más antiguo de la lista para proporcionar más flexibilidad a los movimientos futuros.

Regla de detención. Utilice algún criterio de detención, como un número fijo de iteraciones, una cantidad fija de tiempo del CPU o un número fijo de iteraciones consecutivas que no produzcan ninguna mejoría al mejor valor de la función objetivo. (El último es un criterio particularmente popular.) El proceso también puede detenerse en cualquier iteración donde no existan movimientos factibles en la vecindad local de la solución de prueba actual. Se acepta la mejor solución de prueba que se haya encontrado en cualquier iteración como la solución final.

Este esquema deja algunas preguntas sin contestar.

1. ¿Cuál procedimiento de búsqueda local debe usarse?
2. ¿Cómo debe definir ese procedimiento la *estructura de vecindad* que especifica cuáles soluciones son vecinas inmediatas de cualquier solución de prueba (que se pueden buscar en una misma iteración)?
3. ¿Cuál es la forma en la que los movimientos tabú deben representarse en la lista tabú?
4. ¿Cuál movimiento tabú debe agregarse a la lista tabú en cada iteración?
5. ¿Qué tanto debe mantenerse un movimiento tabú en la lista?
6. ¿Cuál regla de detención debe usarse?

Todos éstos son detalles importantes que deben trabajarse para satisfacer el tipo específico de problema que se aborde, como lo ilustran los siguientes ejemplos. La búsqueda tabú sólo proporciona una estructura general y directrices estratégicas para desarrollar un método heurístico específico para afrontar una situación particular. La selección de estos parámetros es una parte clave en el desarrollo de un método heurístico exitoso.

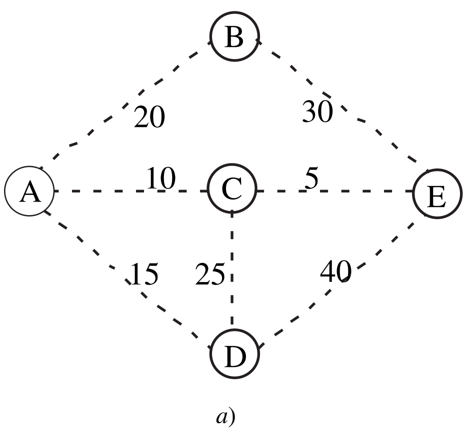
Los siguientes ejemplos ilustran el uso de la búsqueda tabú.

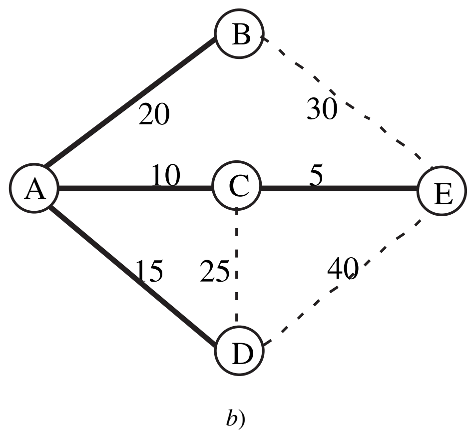
Problema del árbol de expansión mínima con restricciones

En la sección 9.4 se describe el problema del árbol de expansión mínima. En resumen, éste se inicia con una red que tiene sus nodos pero todavía no las ligaduras entre ellos, por lo que el problema es determinar cuáles ligaduras deben insertarse en la red. El objetivo es minimizar el costo (o la longitud) total de las ligaduras insertadas que proporcionarán una ruta entre cada par de nodos. En el caso de una red con n nodos se necesitan $(n-1)$ ligaduras (sin ciclos) para proporcionar una ruta entre cada par de nodos. Una red como ésta es conocida como un *árbol de expansión*.

El lado izquierdo de la figura 13.7 muestra una red con cinco nodos, en la cual las líneas punteadas representan las ligaduras potenciales que podrían insertarse en la red, mientras que el

FIGURA 13.7
a) Datos de un problema de árbol de expansión mínima antes de escoger las ligaduras que deben incluirse en la red y *b*) solución óptima de este problema donde las líneas oscuras representan las ligaduras seleccionadas.





número a continuación de cada línea punteada representa el costo asociado con la inserción de esa ligadura en particular. Por tanto, el problema es determinar las cuatro ligaduras (sin ciclos) que deben insertarse en la red para minimizar el costo total de dichas ligaduras. El lado derecho de la figura presenta el *árbol de expansión mínima* deseado, donde las líneas oscuras representan las ligaduras que se han insertado en la red con un costo total de 50. Esta solución óptima se obtiene con facilidad al aplicar el algoritmo "codicioso" presentado en la sección 9.4.

Para ilustrar el uso de la búsqueda tabú, a continuación se agregará un par de complicaciones a este ejemplo al suponer que también deben observarse las siguientes restricciones en el momento de elegir las ligaduras a incluirse en la red.

Restricción 1: La ligadura AD se puede incluir sólo si también se incluye la ligadura DE.
Restricción 2: Se puede incluir, como máximo, una de las tres ligaduras: AD, CD y AB.

Observe que la solución óptima previa en el lado derecho de la figura 13.7 viola ambas restricciones porque: 1) está incluida la ligadura AD aunque DE no lo está y 2) se incluye tanto AD como AB.

Al imponer estas restricciones, el algoritmo codicioso que se presentó en la sección 9.4 ya no se puede utilizar para encontrar la nueva solución óptima. Para un problema tan pequeño, es probable que esta solución se encuentre con mayor rapidez mediante inspección. Sin embargo, a continuación se verá cómo podría utilizarse la búsqueda tabú, tanto en éste como en problemas mucho más grandes, para buscar una solución óptima.

La manera más fácil de tomar en cuenta estas restricciones es cargar una penalización muy grande por violarlas, como las que se presentan en seguida.

1. Cargar una penalización de 100 si se viola la restricción 1.
2. Cargar una penalización de 100 si se incluyen dos de las tres ligaduras especificadas en la restricción 2. Aumentar esta penalización a 200 si se incluyen las tres ligaduras.

Una penalización de 100 es lo suficientemente grande como para asegurar que las restricciones no se violarán en un árbol de expansión mínima que minimice el costo total, incluyendo las penalizaciones, dado que existen algunas restricciones factibles. Al duplicar la penalización en caso de que la restricción 2 se viole en mayor medida, se proporciona un incentivo para al menos reducir el número de ligaduras que se consideran durante una iteración de la búsqueda tabú.

Existen algunas formas de responder las seis preguntas que se necesitan para especificar la manera en que se realizará una búsqueda tabú. (Vea la lista de preguntas que sigue al esquema de un algoritmo de búsqueda tabú básico.) A continuación se presenta una forma directa de contestar las preguntas.

1. **Procedimiento de búsqueda local:** En cada iteración elija el mejor vecino inmediato de la solución de prueba actual que no esté descartado por la lista tabú.
2. **Estructura de vecindad:** Un vecino inmediato de la solución de prueba actual es aquél al que se llega mediante la adición de una sola ligadura y la eliminación de una de las otras ligaduras dentro del ciclo que se forma por la adición realizada. (La ligadura que se borra debe formar parte de este ciclo para así conservar la forma de un árbol de expansión.)
3. **Forma de los movimientos tabú:** Enumere las ligaduras que no se deben eliminar.
4. **Adición de un movimiento tabú:** En cada iteración, después de elegir la ligadura que debe agregarse a la red, también incorpore esta ligadura a la lista tabú.
5. **Tamaño máximo de la lista tabú:** Dos. Siempre que se agregue un movimiento tabú a una lista llena, elimine el más antiguo de los dos movimientos tabú que ya estaban en la lista. (Como un árbol de expansión del problema bajo consideración sólo incluye cuatro ligaduras, la lista tabú debe ser muy pequeña para proporcionar algo de flexibilidad al elegir la ligadura que debe borrarse en cada iteración.)
6. **Regla de detención:** Detenga el proceso después de tres iteraciones consecutivas sin mejora del mejor valor de la función objetivo. (También debe detenerse en cualquier iteración donde la solución de prueba actual no tenga vecinos inmediatos que no estén descartados por la lista tabú.)

Después de haber especificado estos detalles, se puede proceder a aplicar el algoritmo de búsqueda tabú al ejemplo. Para comenzar, una elección razonable de la solución de prueba inicial

es la solución óptima de la versión no restringida del problema que se muestra en la figura 13.7*b*). Como esta solución viola ambas restricciones (pero con la inclusión de sólo dos de las tres ligaduras especificadas en la restricción 2), es necesario imponer penalizaciones de 100 dos veces. Por lo tanto, el costo total de esta solución es

$$\begin{aligned} \text{Costo} &= 20 + 10 + 5 + 15 + 200 \text{ (penalizaciones de restricción)} \\ &= 250. \end{aligned}$$

Iteración 1. Las tres opciones para agregar una ligadura a la red de la figura 13.7*b*) son BE, CD y DE. Si se eligiera BE, el ciclo formado sería BE-CE-AC-AB, en cuyo caso las tres opciones para borrar una ligadura serían CE, AC y AB. (En este punto todavía no se agrega ninguna ligadura a la lista tabú.) Si se borrara CE, el cambio en el costo sería 30 – 5 = 25 sin cambio en las penalizaciones de restricción, por lo que el costo se incrementaría de 250 a 275. De forma similar, si se borrara AC, el costo total aumentaría de 250 a 250 + (30 – 10) = 270. Sin embargo, si la ligadura borrada fuese AB, los costos de ligadura cambiarían en 30 – 20 = 10 y las penalizaciones de restricción disminuirían de 200 a 100 porque ya no se violaría la restricción 2, por lo cual el costo total ahora sería 50 + 10 + 100 = 160. Estos resultados se resumen en los primeros tres renglones de la tabla 13.1.

En los siguientes dos renglones se resumen los cálculos si se agregara la ligadura CD a la red. En este caso el ciclo creado es CD-AD-AC, de manera que AD y AC son las únicas ligaduras que pueden eliminarse. AC sería una elección particularmente mala porque se seguiría violando la restricción 1 (con penalización de 100), y además se debería cargar una penalización de 200 debido a que se violaría la restricción 2 puesto que las tres ligaduras especificadas en dicha restricción estarían incluidas en la red. Al borrar AD se obtendría el beneficio de satisfacer la restricción 1 y no se incrementaría el grado en el que se viola la restricción 2.

Los últimos tres renglones de la tabla muestran las opciones si se agregara la ligadura DE. El ciclo creado por esta adición sería DE-CE-AC-AD, por lo que CE, AC y AD serían las opciones de eliminación. Las tres satisfarían la restricción 1, pero al borrar AD se cumpliría también con la restricción 2. Si se eliminan por completo las penalizaciones de restricción, el costo total de esta opción sería sólo de 50 + (40 – 15) = 75. Como éste es el costo más pequeño de las ocho opciones disponibles de movimiento hacia un vecino inmediato a partir de la solución de prueba actual, se selecciona esta acción particular al agregar DE y eliminar AD. Esta elección se indica en la parte dedicada a la iteración 1 de la figura 13.8, mientras que el árbol de expansión resultante, con el cual se inicia la iteración 2, se muestra a la derecha.

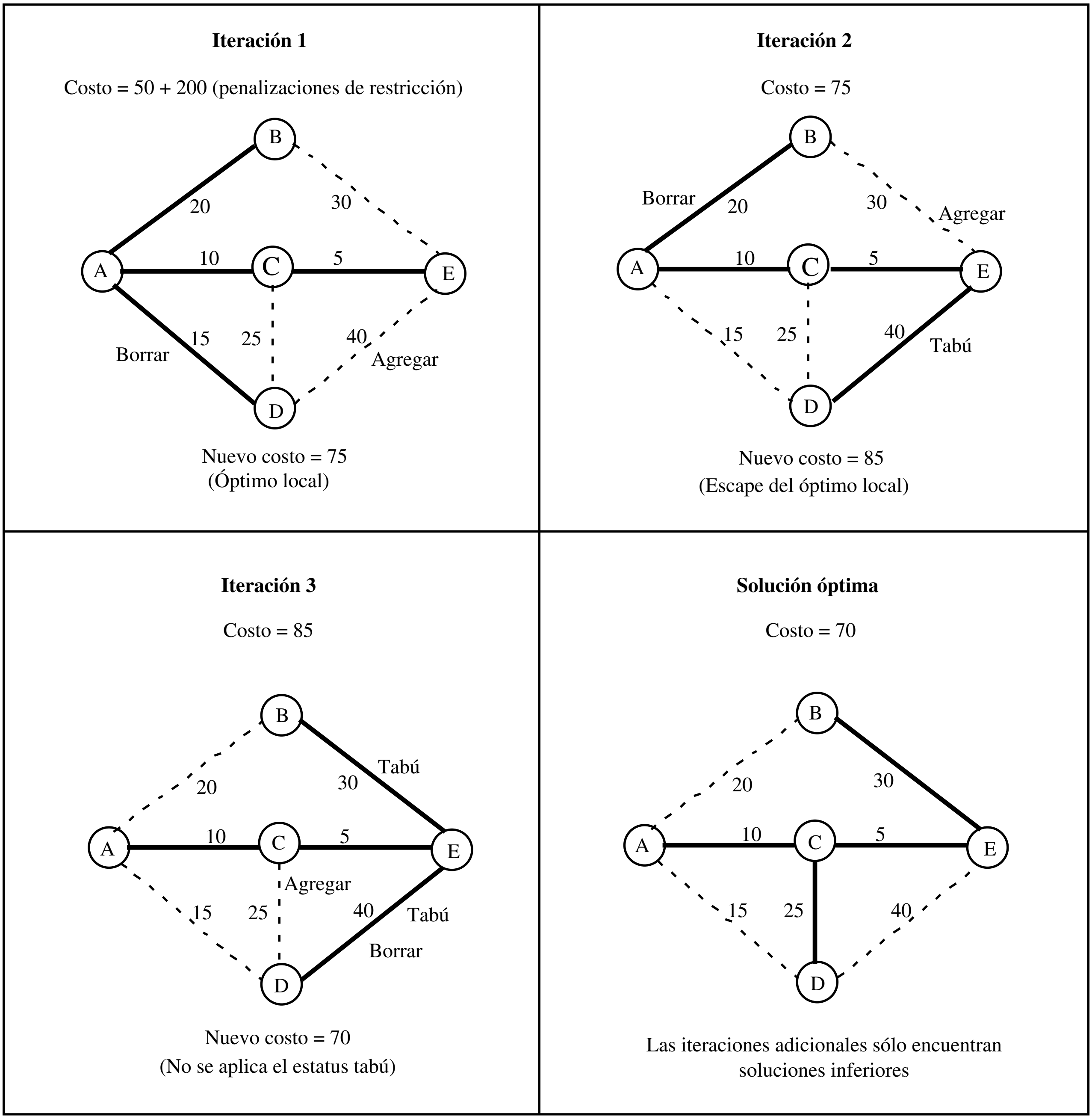
Para completar la iteración, como DE se agregó a la red, se convierte en la primera ligadura colocada en la lista tabú. Esta inclusión evitará que DE sea eliminada en la siguiente iteración y que se regrese a la solución de prueba que inició este proceso.

Para resumir, durante la primera iteración se tomaron las siguientes decisiones.

Se agrega la ligadura DE a la red.
Se elimina la ligadura AD de la red.
Se añade la ligadura DE a la lista tabú.

TABLA 13.1 Opciones para agregar y borrar ligaduras en la iteración 1

Agregar	Borrar	Costo
BE	CE	75 + 200 = 275
BE	AC	70 + 200 = 270
BE	AB	60 + 100 = 160
CD	AD	60 + 100 = 160
CD	AC	65 + 300 = 365
DE	CE	85 + 100 = 185
DE	AC	80 + 100 = 180
DE	AD	75 + 0 = 75 ←Mínimo



■ **FIGURA 13.8**
Aplicación de un algoritmo de búsqueda tabú al problema de árbol de expansión mínima que se presentó en la figura 13.7, después de agregarle dos restricciones.

Iteración 2. La parte superior derecha de la figura 13.8 indica que las decisiones correspondientes que se tomaron durante la iteración 2 son las siguientes:

Se agrega la ligadura BE a la red.
De manera automática, esta ligadura agregada se coloca en la lista tabú.
Se elimina la ligadura AB de la red.

En la tabla 13.2 se resumen los cálculos que condujeron a estas decisiones al encontrar que el movimiento del sexto renglón proporciona el menor costo.

TABLA 13.2 Opciones para agregar y borrar ligaduras en la iteración 2

Agregar	Borrar	Costo
AD	DE*	(Movimiento tabú)
AD	CE	85 + 100 = 185
AD	AC	80 + 100 = 180
BE	CE	100 + 0 = 100
BE	AC	95 + 0 = 95
BE	AB	85 + 0 = 85 ← Mínimo
CD	DE*	60 + 100 = 160
CD	CE	95 + 100 = 195

*Movimiento tabú. Será tomado en cuenta sólo si genera una mejor solución que la mejor solución encontrada con anterioridad.

Los movimientos que se mencionan en los renglones primero y séptimo de la tabla implican la eliminación de DE, lo que está en la lista tabú. Por tanto, estos movimientos se hubieran considerado sólo si hubiesen proporcionado una mejor solución que la mejor solución de prueba encontrada hasta el momento, la cual tiene un costo de 75 dólares. El cálculo en el séptimo renglón muestra que este movimiento no proporcionaría una mejor solución. En el caso del primer renglón ni siquiera es necesario realizar un cálculo porque este movimiento regresaría el proceso a la solución de prueba anterior.

Observe que el movimiento del sexto renglón se realiza aunque genera una nueva solución de prueba con un costo más elevado (85) que el de la solución anterior (75) con la que se inició la iteración 2. Esto significa que la solución de prueba anterior fue un óptimo local porque todos sus vecinos inmediatos (aquellos a los que se puede llegar mediante uno de los movimientos que se mencionan en la tabla 13.2) tienen un costo más alto. Sin embargo, al desplazarse hacia el mejor de los vecinos inmediatos, el proceso tiene la posibilidad de escapar del óptimo local y continuar la búsqueda del óptimo global.

Antes de continuar con la iteración 3 es necesario hacer una observación sobre las formas más avanzadas de búsqueda tabú que se puede usar para seleccionar el mejor vecino inmediato. Algunos métodos más generales de búsqueda tabú pueden cambiar el significado de “el mejor vecino”, lo cual depende de la historia. Éstos pueden utilizar formas adicionales de memoria para apoyar procesos de intensificación y diversificación. Como ya se mencionó, la intensificación enfoca la búsqueda en una región particularmente promisoria en cuanto a soluciones, mientras que la diversificación conduce la búsqueda a nuevas regiones promisorias.

Iteración 3. En la parte inferior izquierda de la figura 13.8 se resumen las decisiones que se tomaron durante la iteración 3.

Se agrega la ligadura CD a la red.
De manera automática, esta ligadura agregada se coloca en la lista tabú.
Se elimina la ligadura DE de la red.

En la tabla 13.3 se muestra que este movimiento conduce al mejor vecino inmediato de la solución de prueba con la que se inició esta iteración.

Una característica interesante de este movimiento es que se realiza aunque sea un movimiento tabú. La razón por la que se hace es que, además de ser el mejor vecino inmediato, también genera una solución que es mejor (con un costo de 70) que la mejor solución de prueba que se encontró con anterioridad (con un costo de 75). Esto permite que el estatus tabú del movimiento se pase por alto. (La búsqueda tabú también puede incorporar una variedad de criterios más avanzados para ignorar el estatus tabú.)

Antes de comenzar la siguiente iteración es necesario hacer otro ajuste a la lista tabú.

Se borra la ligadura DE de la lista tabú.

TABLA 13.3 Opciones para agregar y borrar ligaduras en la iteración 3

Agregar	Borrar	Costo
AB	BE*	(Movimiento tabú)
AB	CE	100 + 0 = 100
AB	AC	95 + 0 = 95
AD	DE*	60 + 100 = 160
AD	CE	95 + 0 = 95
AD	AC	90 + 0 = 90
CD	DE*	70 + 0 = 70 ← Mínimo
CD	CE	105 + 0 = 105

*Movimiento tabú. Será tomado en cuenta sólo si genera una mejor solución que la mejor solución de prueba encontrada anteriormente.

Este paso se da por dos razones. Primero, la lista tabú consiste en ligaduras que por lo general no se deben borrar de la red durante la iteración actual (con la excepción que se mencionó), pero DE ya no está en la red. Segundo, como el tamaño de la lista tabú se estableció en dos y hay otras dos ligaduras (BE y CD) que se han agregado a la lista más recientemente, de cualquier manera DE tendría que haberse borrado de manera automática en este punto.

Continuación. La solución de prueba actual que se muestra en la parte inferior derecha de la figura 13.8 es, en realidad, la solución óptima (el óptimo global) del problema. Sin embargo, el algoritmo de búsqueda tabú no tiene forma de saberlo, por lo que continuará con el proceso de búsqueda. La iteración 4 comenzaría con esta solución de prueba y con las ligaduras BE y CD en la lista tabú. Después de completar esta iteración y dos más, el algoritmo terminaría porque después de tres iteraciones consecutivas no habría mejoría sobre el mejor valor previo de la función objetivo (un costo de 70).

Con un algoritmo de búsqueda tabú bien diseñado, es probable que la mejor solución de prueba que se encuentre después de haber realizado un número modesto de iteraciones sea una buena solución factible. Podría incluso ser una solución óptima, pero no existe garantía de ello. Seleccionar una regla de detención que proporcione una corrida relativamente de largo plazo del algoritmo incrementa las posibilidades de alcanzar el óptimo global.

Después de haber explicado la aplicación de este algoritmo mediante su ejecución en un ejemplo pequeño, ahora se aplicará un algoritmo de búsqueda tabú similar al ejemplo del agente viajero que se presentó en la sección 13.1.

Ejemplo del problema del agente viajero

Existen algunos paralelismos cercanos entre un problema de árbol de expansión mínima y el del agente viajero. En ambos casos, el problema consiste en elegir las ligaduras que deben incluirse en la solución. (Recuerde que una solución de un problema del agente viajero puede describirse como la secuencia de ligaduras por las que pasa el agente a lo largo de su recorrido por las ciudades.) En ambos casos, el objetivo es minimizar el costo o la distancia total asociada con el número fijo de ligaduras que se incluye en la solución. En ambos casos existe un procedimiento de búsqueda local intuitivo que implica la adición y eliminación de ligaduras en la solución de prueba actual para obtener la nueva solución de prueba.

En el caso de los problemas de árbol de expansión mínima, el procedimiento de búsqueda local que se describió en la subsección anterior implica la adición y la eliminación de *una sola* ligadura en cada iteración. El procedimiento correspondiente que se explicó en la sección 13.1 del problema del agente viajero implica la utilización de *subviajes inversos* para agregar y borrar un *par* de ligaduras en cada iteración.

Debido a los paralelismos cercanos entre estos dos tipos de problemas, el diseño de un algoritmo de búsqueda tabú para problemas del agente viajero puede ser muy similar al que se acaba de describir para el ejemplo del problema de árbol de expansión mínima. En particular, si se usa el

esquema del algoritmo de búsqueda tabú básico que se presentó antes, es posible responder de una manera similar las seis preguntas que siguen al esquema, como se indica a continuación.

1. **Algoritmo de búsqueda local:** En cada iteración seleccione el mejor vecino inmediato de la solución de prueba actual que no esté descartado por la lista tabú.
2. **Estructura de vecindad:** Un vecino inmediato de la solución de prueba actual es aquel al que se llega por medio de un *subviaje inverso*, como se describe en la sección 13.1 y se ilustra en la figura 13.5. Tal inversión requiere la adición de dos ligaduras y la eliminación de otras dos ligaduras de la solución de prueba actual. (Se descarta un subviaje inverso que sólo invierta la dirección del viaje que proporciona la solución de prueba actual.)
3. **Forma de los movimientos tabú:** Enumere las ligaduras de forma que un subviaje inverso sea tabú si *las dos* ligaduras que se eliminan por esta inversión se encuentran en la lista. (Ello evitará que se regrese con rapidez a la solución de prueba anterior.)
4. **Adición de un movimiento tabú:** En cada iteración, después de elegir las dos ligaduras que deben agregarse a la solución de prueba actual, también incorpore estas dos ligaduras a la lista tabú.
5. **Tamaño máximo de la lista tabú:** Cuatro (dos de cada una de las dos iteraciones más recientes). Siempre que se agregue un par de ligaduras a una lista llena, elimine las dos ligaduras que han permanecido por más tiempo en la lista.
6. **Regla de detención:** Detenga el proceso después de tres iteraciones consecutivas sin mejora del mejor valor de la función objetivo. (También debe detenerse en cualquier iteración donde la solución de prueba actual no tenga vecinos inmediatos que no estén descartados por la lista tabú.)

Para aplicar este algoritmo de búsqueda tabú al ejemplo (vea la figura 13.4), se comenzará con la misma solución de prueba inicial, 1-2-3-4-5-6-7-1, como en la sección 13.1. Recuerde cómo al empezar el algoritmo del subviaje inverso (un algoritmo de mejora local), con esta solución de prueba inicial se llega en dos iteraciones (vea las figuras 13.5 y 13.6) a un óptimo local en 1-2-4-6-5-3-7-1, en cuyo punto el algoritmo se detiene. Con la excepción de que se crea una lista tabú, el algoritmo de búsqueda tabú se inicia exactamente del mismo modo, como se resume a continuación.

Solución de prueba inicial: 1-2-3-4-5-6-7-1 Distancia = 69
Lista tabú: En blanco.

Iteración 1: Se elige invertir 3-4 (vea la figura 13.5).
Ligaduras borradas: 2-3 y 4-5
Ligaduras agregadas: 2-4 y 3-5
Lista tabú: Ligaduras 2-4 y 3-5.
Nueva solución de prueba: 1-2-4-3-5-6-7-1 Distancia = 65

Iteración 2: Se elige invertir 3-5-6 (vea la figura 13.6).
Ligaduras borradas: 4-3 y 6-7 (no están en la lista tabú)
Ligaduras agregadas: 4-6 y 3-7
Lista tabú: Ligaduras 2-4, 3-5, 4-6 y 3-7
Nueva solución de prueba: 1-2-4-6-5-3-7-1 Distancia = 64

Sin embargo, en lugar de terminar, el algoritmo de búsqueda tabú escapa de este óptimo local (que se muestra en el lado derecho de la figura 13.6 y en el lado izquierdo de la figura 13.9) pues se traslada hacia el mejor vecino inmediato de la solución de prueba actual aunque su distancia sea mayor. Debido a la disponibilidad limitada de ligaduras entre pares de nodos (ciudades) de la figura 13.4, la solución de prueba actual sólo tiene dos vecinos inmediatos, los cuales de describen a continuación.

Si se invierte 6-5-3: 1-2-4-3-5-6-7-1 Distancia = 65
Si se invierte 3-7: 1-2-4-6-5-7-3-1 Distancia = 66

(Se descarta invertir 2-4-6-5-3-7 para obtener 1-7-3-5-6-4-2-1 porque es exactamente el mismo viaje pero en la dirección opuesta.) Sin embargo, se debe descartar el primero de estos vecinos inmediatos porque requeriría eliminar las ligaduras 4-6 y 3-7, lo cual está prohibido porque *las dos* ligaduras se encuentran en la lista tabú. (Este movimiento se podría permitir si implicara alguna mejoría en la mejor solución de prueba que se encontró hasta ahora, pero no es así.) Al descartar este vecino inmediato simplemente se evita regresar a la solución de prueba anterior. Entonces, por no existir otro, se selecciona el segundo de los vecinos inmediatos como la próxima solución de prueba, como se resume a continuación.

Iteración 3: Se elige invertir 3-7 (vea la figura 13.9).
Ligaduras borradas: 5-3 y 7-1
Ligaduras agregadas: 5-7 y 3-1
Lista tabú: 4-6, 3-7, 5-7 y 3-1
(2-4 y 3-5 se eliminan de la lista)
Nueva solución de prueba: 1-2-4-6-5-7-3-1 Distancia = 66

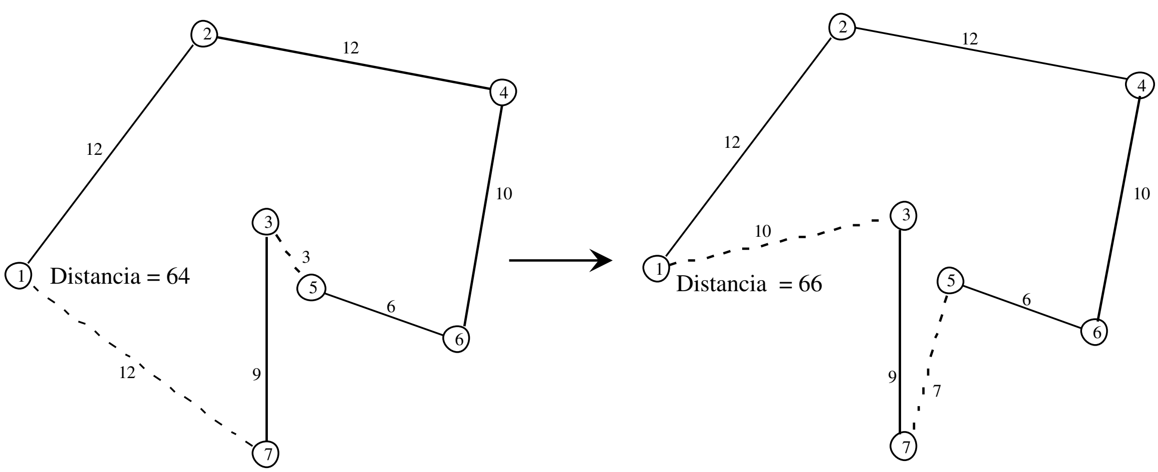
La inversión del subviaje de esta iteración se puede ver en la figura 13.9 en la cual las líneas punteadas muestran las ligaduras borradas (a la izquierda) y las agregadas (a la derecha) para obtener la nueva solución de prueba. Observe que una de las ligaduras borradas es 5-3 aun cuando estaba en la lista tabú al final de la iteración 2. Esto es correcto porque un subviaje inverso es tabú sólo si *las dos* ligaduras borradas están en la lista tabú. También observe que la lista tabú actualizada al final de la iteración 3 ya no contiene las dos ligaduras que habían permanecido por más tiempo en dicha lista (las que se agregaron durante la iteración 1), puesto que el tamaño máximo de la lista tabú se estableció en cuatro.

La nueva solución de prueba tiene los cuatro vecinos inmediatos que se describen a continuación.

Invertir 2-4-6-5-7: 1-7-5-6-4-2-3-1 Distancia = 65
Invertir 6-5: 1-2-4-5-6-7-3-1 Distancia = 69
Invertir 5-7: 1-2-4-6-5-7-3-1 Distancia = 63
Invertir 7-3: 1-2-4-6-5-3-7-1 Distancia = 64

Sin embargo, el segundo de estos vecinos inmediatos es tabú porque *las dos* ligaduras borradas (4-6 y 5-7) están en la lista tabú. Por la misma razón, el cuarto vecino inmediato (que es la solución de prueba anterior) también es tabú. Por lo tanto, las únicas opciones viables son el primero y el tercero de los vecinos inmediatos. Como el último de éstos tiene la distancia más corta, se convierte en la próxima solución de prueba, como se resume a continuación.

■ **FIGURA 13.9**
Inversión del subviaje 3-7 en la iteración 3 que conduce desde la solución de prueba de la izquierda hasta la nueva solución de prueba de la derecha.



Iteración 4: Se elige invertir 5-7 (vea la figura 13.10).
Ligaduras borradas: 6-5 y 7-3
Ligaduras agregadas: 6-7 y 5-3
Lista tabú: 5-7, 3-1, 6-7 y 5-3
(4-6 y 3-7 se eliminan de la lista)
Nueva solución de prueba: 1-2-4-6-7-5-3-1 Distancia = 63

En la figura 13.10 se muestra esta inversión del subviaje. El viaje de la nueva solución de prueba a la derecha tiene una distancia de sólo 63, que es menor que cualquiera de las soluciones anteriores. En realidad, esta nueva solución resulta ser la solución óptima.

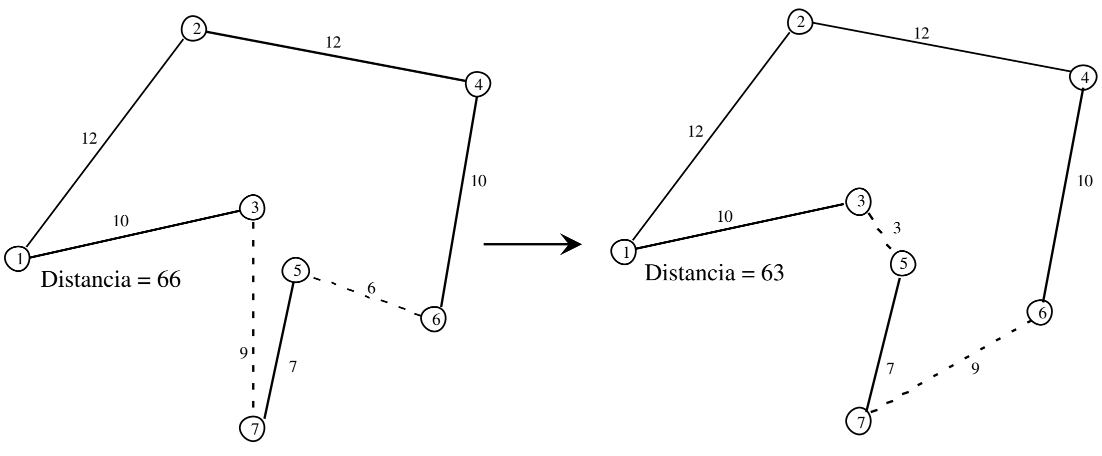
Si no se sabe esto, el algoritmo de búsqueda tabú intentará ejecutar más iteraciones. Sin embargo, el único vecino inmediato de la solución de prueba actual es la solución de prueba que se obtuvo en la iteración anterior. Esto requeriría eliminar las ligaduras 6-7 y 5-3, que se encuentran en la lista tabú, de manera que es imposible regresar a la solución de prueba anterior. Como ya no existe otro vecino inmediato disponible, la regla de detención da por finalizado el algoritmo en este punto con 1-2-4-6-7-5-3-1 (la mejor de las soluciones de prueba) como la solución final. Aunque no hay garantía de que la solución final del algoritmo sea una solución óptima, resulta afortunado que en este caso la solución haya resultado un óptimo global.

En el área de metaheurísticas del IOR Tutorial se incluye una rutina para aplicar este algoritmo de búsqueda tabú particular a otros pequeños problemas del agente viajero.

Este algoritmo en específico es sólo un ejemplo de un algoritmo de búsqueda tabú posible para manejar problemas del agente viajero. Varios detalles del algoritmo pueden modificarse en diversas formas razonables. Por ejemplo, por lo general el método no se detiene cuando todos los movimientos disponibles están prohibidos por la lista tabú, sino que selecciona el movimiento "menos tabú". También, una característica importante de los métodos de búsqueda tabú generales es el uso de vecindades múltiples, que descansa en vecindades básicas cuando comienza el proceso para después incluir vecindades más avanzadas cuando la tasa a la que se encuentran soluciones mejoradas disminuye. El elemento adicional más significativo de la búsqueda tabú es el uso de estrategias de intensificación y diversificación, que se mencionó con anterioridad. Sin embargo, el esquema básico general de un enfoque de búsqueda tabú con "memoria a corto plazo", en lo general, se conserva igual que como se ha ilustrado aquí.

Los dos ejemplos considerados en esta sección caen en la categoría de problemas de optimización combinatoria que involucran redes. Ésta es un área de aplicación particularmente común de los algoritmos de búsqueda tabú. El esquema general de estos algoritmos incorpora los principios que se presentaron en esta sección, pero los detalles se trabajan para ajustarse a la estructura del problema específico bajo consideración.

FIGURA 13.10
Inversión del subviaje 5-7 en la iteración 4 que conduce desde la solución de prueba de la izquierda hasta la nueva solución de prueba de la derecha (que resulta ser la solución óptima).



13.3 TEMPLADO SIMULADO

El templado simulado es otra metaheurística que se utiliza con amplitud que permite al proceso de búsqueda escapar de un óptimo local. Para compararlo y contrastarlo mejor con la búsqueda tabú, se aplicará al mismo problema del agente viajero antes de regresar al ejemplo de programación no lineal que se explicó en la sección 13.1. Pero primero se examinarán los conceptos básicos del templado simulado.

Conceptos básicos

En la figura 13.1 de la misma sección se introdujo el concepto de que la localización del óptimo global de un problema complicado de maximización es análogo a determinar cuál de todos las montañas existentes es la más alta y después escalar hasta la cumbre de esta montaña. Desafortunadamente, un proceso de búsqueda matemática no tiene el beneficio de una visión aguda que pudiera ayudar a rastrear una cumbre a la distancia. En su lugar, el proceso consiste en algo parecido a escalar en medio de una niebla densa donde la única pista de la dirección que debe tomarse es qué tan pronunciada es la pendiente ascendente o descendente del siguiente paso.

Un enfoque, adoptado por la búsqueda tabú, es escalar la pendiente actual en la dirección más empinada hasta alcanzar ese pico particular y después descender lentamente para buscar otra cuesta que escalar. La desventaja es que se emplea demasiado tiempo (iteraciones) en escalar cada cuesta que se encuentra en lugar de buscar la más alta.

Por otro lado, el enfoque que se utiliza en el templado simulado es concentrarse de manera principal en la búsqueda de la cuesta más alta. Como el pico más alto puede estar en cualquier sitio dentro de la región factible, se le otorga la mayor importancia a dar pasos en direcciones aleatorias (excepto por el rechazo de algunos, pero no de todos, los pasos que descienden en lugar de ascender) con la intención de explorar tanta región factible como sea posible. Como la mayoría de los pasos aceptados son ascendentes, en forma gradual la búsqueda gravitará hacia aquellas partes de la región factible que contienen las cumbres más altas. Por lo tanto, de manera gradual, el proceso de búsqueda incrementa el hincapié en el ascenso al rechazar una proporción creciente de pasos que descienden. Si se da el tiempo suficiente, con frecuencia el proceso encontrará y ascenderá hasta la cuesta más alta.

De manera más específica, cada iteración del proceso de búsqueda de templado simulado se mueve de la solución de prueba actual a un vecino inmediato en la vecindad local de esta solución, igual que en la búsqueda tabú. Sin embargo, la diferencia con ésta reside en cómo se selecciona un vecino inmediato para ser la próxima solución de prueba. Sea

$Z_c =$ valor de la función objetivo de la solución de prueba *actual*,
$Z_n =$ valor de la función objetivo del candidato actual a ser la siguiente solución de prueba,
$T =$ un parámetro que mide la tendencia a aceptar el candidato actual para ser la próxima solución de prueba si este candidato no es una mejora sobre la solución de prueba actual.

La regla para seleccionar cuál vecino inmediato será la próxima solución de prueba es la siguiente.

Regla de selección del movimiento: Entre todos los vecinos inmediatos de la solución de prueba actual, seleccione uno de manera aleatoria para convertirse en el candidato actual a ser la próxima solución de prueba. Si se supone que el objetivo es la *maximización* de la función objetivo, acepte o rechace este candidato para ser la próxima solución de prueba como sigue:

Si $Z_n \geq Z_c$, siempre acepte este candidato.
Si $Z_n < Z_c$, acepte el candidato con la siguiente probabilidad:

$$\text{Prob}\{\text{aceptación}\} = e^x \text{ donde } x = \frac{Z_n - Z_c}{T}$$

(Si el objetivo es de *minimización*, se deben invertir Z_n y Z_c en las fórmulas anteriores.) Si el candidato es rechazado repita el proceso con un vecino inmediato de la solución de prueba actual seleccionado de manera aleatoria. (Si ya no existen vecinos inmediatos restantes, el algoritmo termina.)

TABLA 13.4 Algunas probabilidades de muestra de que la regla de selección del movimiento acepte un paso descendente cuando el objetivo es maximizar

$x = \frac{Z_n - Z_c}{T}$	Prob{aceptación} = e^x
−0.01	0.990
−0.1	0.905
−0.25	0.779
−0.5	0.607
−1	0.368
−2	0.135
−3	0.050
−4	0.018
−5	0.007

Por tanto, si el candidato actual bajo consideración es mejor que la solución de prueba actual, éste siempre se acepta como la próxima solución de prueba. Si es peor, la probabilidad de aceptación depende de qué tan peor es (y del tamaño de T). En la tabla 13.4 se presenta una muestra de estos valores de probabilidad, que van desde una muy alta cuando el candidato actual es sólo un poco peor (en relación con T) que la solución de prueba actual hasta una probabilidad en extremo pequeña cuando es mucho peor. En otras palabras, la regla de selección del movimiento por lo general aceptará un paso que sólo es un poco descendente, pero será muy difícil que acepte un paso descendente muy pronunciado. Si se inicia con un valor relativamente grande de T (como lo hace el templado simulado) la probabilidad de aceptación correspondiente será alta, lo que permite a la búsqueda proceder en direcciones casi aleatorias. Si se disminuye de manera gradual el valor de T a medida que la búsqueda continúa (como lo hace el templado simulado) la probabilidad de aceptación disminuye de un modo paulatino, lo cual aumenta la importancia de ascender la mayor parte del tiempo. Así, la elección de los valores de T a través del tiempo controla el grado de aleatoriedad del proceso para permitir pasos descendentes. Este componente aleatorio, que no está presente en la búsqueda tabú básica, proporciona más flexibilidad de movimiento hacia otra parte de la región factible con la esperanza de encontrar una cumbre más alta.

El método usual de implantar la regla de selección del movimiento para determinar si se aceptará un paso descendente particular es comparar un **número aleatorio** entre 0 y 1 con la probabilidad de aceptación. Ese número aleatorio puede considerarse como una observación aleatoria en una distribución uniforme entre 0 y 1. (A lo largo de este capítulo todas las referencias a números aleatorios tendrán relación con ese tipo de números aleatorios.) Existen varios métodos para generar estos números aleatorios (como se describirá en la sección 20.3). Por ejemplo, la función de Excel ALEATORIO() los genera cuando se le solicita. (También, al inicio de la sección de problemas se describe cómo pueden usarse los dígitos aleatorios de la tabla 20.3 para obtener los números aleatorios que se necesitarán para resolver algunos de los problemas de tarea.)

Si *número aleatorio* $<$ Prob{aceptación}, se acepta un paso descendente.
En otro caso, se rechaza el paso.

¿Por qué el templado simulado usa la fórmula particular para Prob{aceptación} que especifica la regla de selección del movimiento? La razón es que el templado simulado se basa en la analogía con un *proceso de templado físico*. En principio, este proceso implica fundir un metal o cristal a una temperatura alta para después enfriar lentamente la sustancia hasta que alcance un estado estable de energía con propiedades físicas deseables. A una temperatura dada T durante este proceso, el nivel de energía de los átomos de la sustancia fluctúa pero tiende a disminuir. Un modelo matemático de cómo fluctúa el nivel de energía supone que los cambios ocurren de manera aleatoria excepto que sólo se aceptan algunos de los incrementos. En particular, la probabilidad de aceptar un incremento cuando la temperatura es T tiene la misma forma que para Prob{aceptación} en la regla de selección del movimiento del templado simulado.

La analogía con un problema de optimización en la forma de minimización es que el nivel de energía de la sustancia en el estado actual del sistema corresponde al valor de la función objetivo

con la solución factible actual del problema. El objetivo de que la sustancia alcance un nivel estable con un nivel de energía tan pequeño como sea posible corresponde al problema de llegar a una solución factible con un valor de la función objetivo tan pequeño como sea posible.

Igual que en un proceso físico de templado, una cuestión clave cuando se diseña un algoritmo de templado simulado de un problema de optimización es seleccionar un **programa de temperatura** apropiado para ser usado. (Debido a la analogía con el templado físico, en el algoritmo de templado simulado se hará referencia a T como la temperatura.) Este programa debe especificar el valor inicial de T, relativamente grande, así como los valores subsecuentes que se reducen en forma progresiva. También debe precisar cuántos movimientos (iteraciones) se deben hacer para cada valor de T. La selección de estos parámetros para ajustarse al problema bajo consideración es un factor clave de la eficacia del algoritmo. Se puede utilizar cierta experimentación preliminar para guiar la selección de los parámetros del algoritmo. Después se especificará un programa de temperatura específico que parece razonable para los dos ejemplos que se consideran en esta sección, pero también se podrían aplicar muchos otros.

Con esta base, ahora es posible proporcionar un esquema del algoritmo de templado simulado básico.

Bosquejo de un algoritmo de templado simulado básico

Paso inicial. Comience con una solución de prueba inicial.

Iteración. Utilice la *regla de selección del movimiento* para elegir la próxima solución de prueba. (Si ninguno de los vecinos inmediatos de la solución de prueba actual se acepta, el algoritmo termina.)

Verificación del programa de temperatura. Después de realizar el número deseado de iteraciones en el valor actual de T, disminuya T hasta el siguiente valor en el programa de temperatura y reinicie las iteraciones en dicho valor.

Regla de detención. Cuando ya se ha realizado el número de iteraciones deseado en el valor más pequeño de T en el programa de temperatura (o cuando no se acepta ninguno de los vecinos inmediatos de la solución de prueba actual), se detiene el algoritmo. Acepte la mejor solución de prueba que encontró en cualquier iteración (incluso de los valores grandes de T) como la solución final.

Antes de aplicar este algoritmo a cualquier problema particular, es necesario trabajar cierto número de detalles para ajustarse a la estructura del problema.

1. ¿Cómo debe seleccionarse la solución de prueba inicial?
2. ¿Cuál es la *estructura de vecindad* que especifica las soluciones que son vecinos inmediatos de cualquier solución de prueba actual (a los que se puede llegar en una sola iteración)?
3. ¿Qué dispositivo debe usarse en la regla de selección del movimiento para elegir en forma *aleatoria* uno de los vecinos inmediatos de la solución de prueba actual para convertirlo en el candidato actual a ser la próxima solución de prueba?
4. ¿Cuál es el programa de temperatura apropiado?

Se ilustrarán algunas formas razonables de contestar estas preguntas en el contexto de la aplicación del algoritmo de templado simulado a los siguientes dos ejemplos.

Ejemplo del problema del agente viajero

De nuevo se considerará el problema particular del agente viajero que se introdujo en la sección 13.1 y se desplegó en la figura 13.4.

En el área de metaheurísticas del IOR Tutorial se incluye una rutina para aplicar el algoritmo de templado simulado básico a pequeños problemas del agente viajero como este ejemplo. Esta rutina responde las cuatro preguntas de la siguiente manera:

1. **Solución de prueba inicial:** Se puede introducir cualquier solución factible (secuencia de ciudades en el viaje), quizá al generar en forma aleatoria la secuencia, pero resulta útil introducir

alguna que parezca una buena solución factible. Por ejemplo, la solución factible 1-2-3-4-5-6-7-1 es una elección razonable.

2. **Estructura de vecindad:** Un vecino inmediato a la solución de prueba actual es alguno al que se pueda llegar mediante un *subviaje inverso*, como se describe en la sección 13.1 y se ilustra en la figura 13.5. (No obstante, se descarta el subviaje inverso que sólo invierte la dirección del viaje que proporciona la solución de prueba actual.)
3. **Selección aleatoria de un vecino inmediato:** La selección de un subviaje que debe invertirse requiere la selección del espacio en la secuencia actual de ciudades donde actualmente se inicia el subviaje y después el espacio donde ahora termina. El espacio de inicio puede ser cualquiera, excepto el primero y el último (reservados para la ciudad de residencia), y el espacio contiguo al último. El espacio de terminación debe estar en cualquier punto después del espacio inicial, excepto en el último. (Un inicio en el segundo espacio y terminación en el espacio contiguo al último también se descarta puesto que éste sólo invertiría la dirección del viaje.) Como se ilustrará más adelante, se utilizan números aleatorios para obtener probabilidades iguales de selección para cualquiera de los espacios de inicio y de terminación elegibles. Si esta selección de los espacios de inicio y terminación resulta ser no factible (porque las ligaduras necesarias para completar el subviaje inverso no están disponibles), el proceso se repite hasta que se encuentra una selección factible.
4. **Programa de temperatura:** Se realizan cinco iteraciones a cada uno de los cinco valores de T (T_1, T_2, T_3, T_4, T_5), donde

$T_1 = 0.2Z_c$ cuando Z_c es el valor de la función objetivo de la solución de prueba inicial,
$T_2 = 0.5T_1$,
$T_3 = 0.5T_2$,
$T_4 = 0.5T_3$,
$T_5 = 0.5T_4$,

Este programa de temperatura particular sólo es ilustrativo de lo que se podría utilizar. $T_1 = 0.2Z_c$ es una elección razonable porque T_1 debe tender a ser bastante grande en comparación con los valores típicos de $|Z_n - Z_c|$, lo que estimulará una búsqueda casi aleatoria a través de la región factible para encontrar dónde debe enfocarse la búsqueda. Sin embargo, para el momento en que T se reduce hasta T_5, casi no se aceptan movimientos sin mejora, por lo que el acento estará en mejorar el valor de la función objetivo.

Cuando se enfrenten problemas más grandes, es probable que se realicen más de cinco iteraciones de cada valor de T. Aún más, es probable que los valores de T se reduzcan con mayor lentitud respecto del programa de temperatura descrito anteriormente.

A continuación se explicará cómo se hace la selección aleatoria de un vecino inmediato. Suponga que se tiene la solución de prueba inicial 1-2-3-4-5-6-7-1 en el ejemplo.

Solución de prueba inicial: 1-2-3-4-5-6-7-1 $\quad Z_c = 69 \quad T_1 = 0.2Z_c = 13.8$

El subviaje que debe invertirse puede comenzar en cualquier lugar entre el segundo espacio (que en la actualidad designa a la ciudad 2) y el sexto espacio (que en la actualidad designa a la ciudad 6). A cada uno de estos cinco espacios se le puede otorgar una misma probabilidad mediante la asignación de los siguientes valores de un número aleatorio entre 0 y 1 para elegir el espacio que se indica a continuación.

0.0000-0.1999: El subviaje se inicia en el espacio 2.
0.2000-0.3999: El subviaje se inicia en el espacio 3.
0.4000-0.5999: El subviaje se inicia en el espacio 4.
0.6000-0.7999: El subviaje se inicia en el espacio 5.
0.8000-0.9999: El subviaje se inicia en el espacio 6.

Suponga que el número aleatorio generado resultó ser 0.2779.

0.2779: Se elige un subviaje que se inicia en el espacio 3.

Al comenzar en el espacio 3, el subviaje que se invertirá necesita terminar en algún lugar entre los espacios 4 y 7. A estos cuatro espacios se les proporcionan probabilidades iguales mediante la siguiente correspondencia con un número aleatorio.

0.0000-0.2499: El subviaje se inicia en el espacio 4.
0.2500-0.4999: El subviaje se inicia en el espacio 5.
0.5000-0.7499: El subviaje se inicia en el espacio 6.
0.7500-0.9999: El subviaje se inicia en el espacio 7.

Suponga que el número aleatorio generado para este propósito resultó ser 0.0461.

0.0461: Se elige el final del subviaje en el espacio 4.

Como los espacios 3 y 4 en la actualidad designan que las ciudades 3 y 4 deben ser la tercera y cuarta ciudades visitadas en el viaje, el subviaje entre las ciudades 3 y 4 se invertirá.

Se invierte 3-4 (vea la figura 13.5): 1-2-4-3-5-6-7-1 $Z_n = 65$

Este vecino inmediato de la solución actual (inicial) se convierte en el candidato actual a ser la mejor solución de prueba. Como

$$Z_n = 65 < Z_c = 69,$$

este candidato es mejor que la solución de prueba actual (recuerde que el objetivo aquí es *minimizar* la distancia total del viaje), por lo cual este candidato se acepta de manera automática para ser la siguiente solución de prueba.

Esta selección de un subviaje inverso fue afortunada porque condujo a una solución factible. Esto no sucede siempre en los problemas del agente viajero, como en el ejemplo donde ciertos pares de ciudades no están conectados de manera directa por una ligadura. Por ejemplo, si los números aleatorios hubieran ordenado la inversión de 2-3-4-5 para obtener el viaje 1-5-4-3-2-6-7-1, la figura 13.4 muestra que ésta es una solución no factible porque no existe liga entre las ciudades 1 y 5 así como entre las ciudades 2 y 6. Cuando se presenta esta situación, es necesario generar nuevos pares de números aleatorios hasta que se obtenga una solución factible. (También se puede construir un procedimiento más complejo para generar números aleatorios sólo para las ligaduras relevantes.)

Para ilustrar un caso donde el candidato actual a ser la próxima solución de prueba es peor que la solución de prueba actual, suponga que la segunda iteración resultó en la inversión de 3-5-6 (como en la figura 13.6) para obtener 1-2-4-6-5-3-7-1, que tiene una distancia total de 64. Después suponga que la tercera iteración comienza al invertir 3-7 (como en la figura 13.9) para obtener 1-2-4-6-5-7-3-1 (que tiene una distancia total de 66) como el candidato actual a ser la próxima solución de prueba. Como 1-2-4-6-5-3-7-1 (con una distancia total de 64) es la solución de prueba actual de la iteración 3, ahora se tiene

$$Z_c = 64, \qquad Z_n = 66, \qquad T_1 = 13.8.$$

Por lo tanto, como el objetivo aquí es *minimizar*, la probabilidad de aceptación de 1-2-4-6-5-7-3-1 como la próxima solución de prueba es

$$\begin{aligned} \text{Prob}\{\text{aceptación}\} &= e^{(Z_c - Z_n)/T_1} \\ &= e^{-2/13.8} \\ &= 0.865. \end{aligned}$$

Si el siguiente número aleatorio que se genera es menor que 0.865, este candidato a solución se debe aceptar como la siguiente solución de prueba. En caso contrario, se debe rechazar.

En la tabla 13.5 se muestran los resultados de la utilización del IOR Tutorial para aplicar por completo el algoritmo de templado simulado a este problema. Observe que las iteraciones 14 y 16 empatan al encontrar la mejor solución de prueba, 1-3-5-7-6-4-2-1 (que resulta ser la solución óptima junto con el viaje equivalente en la dirección inversa, 1-2-4-6-7-5-3-1), de manera que

TABLA 13.5 Aplicación del algoritmo de templado simulado del IOR Tutorial al ejemplo del problema del agente viajero

Distancia	*T*	Solución de prueba obtenida	Distancia
0		1-2-3-4-5-6-7-1	69
1	13.8	1-3-2-4-5-6-7-1	68
2	13.8	1-2-3-4-5-6-7-1	69
3	13.8	1-3-2-4-5-6-7-1	68
4	13.8	1-3-2-4-6-5-7-1	65
5	13.8	1-2-3-4-6-5-7-1	66
6	6.9	1-2-3-4-5-6-7-1	69
7	6.9	1-3-2-4-5-6-7-1	68
8	6.9	1-2-3-4-5-6-7-1	69
9	6.9	1-2-3-5-4-6-7-1	65
10	6.9	1-2-3-4-5-6-7-1	69
11	3.45	1-2-3-4-6-5-7-1	66
12	3.45	1-3-2-4-6-5-7-1	65
13	3.45	1-3-7-5-6-4-2-1	66
14	3.45	1-3-5-7-6-4-2-1	63 ← Mínimo
15	3.45	1-3-7-5-6-4-2-1	66
16	1.725	1-3-5-7-6-4-2-1	63 ← Mínimo
17	1.725	1-3-7-5-6-4-2-1	66
18	1.725	1-3-2-4-6-5-7-1	65
19	1.725	1-2-3-4-6-5-7-1	66
20	1.725	1-3-2-4-6-5-7-1	65
21	0.8625	1-3-7-5-6-4-2-1	66
22	0.8625	1-3-2-4-6-5-7-1	65
23	0.8625	1-2-3-4-6-5-7-1	66
24	0.8625	1-3-2-4-6-5-7-1	65
25	0.8625	1-3-7-5-6-4-2-1	66

esta solución se acepta como la solución final. El lector interesado podría encontrar muy útil la aplicación de este software al mismo problema una vez más. Dada la aleatoriedad construida en el algoritmo, la secuencia de soluciones de prueba que se obtenga será diferente cada vez. Debido a esta característica, los practicantes reaplican varias veces un algoritmo de templado simulado al mismo problema para incrementar la oportunidad de encontrar una solución óptima. (En el problema 13.3-2 se pide hacerlo para este mismo ejemplo.) La solución de prueba inicial también puede cambiarse cada vez para facilitar una exploración más profunda de toda la región factible.

Si desea ver **otro ejemplo** de cómo se utilizan los números aleatorios para realizar una iteración del algoritmo de templado simulado básico de un problema del agente viajero, en la sección Worked Examples del sitio en internet de este libro se proporciona uno de ellos.

Antes de continuar con el ejemplo siguiente, se debe hacer una pausa en este punto para mencionar un par de formas en las que las avanzadas características de la búsqueda tabú se pueden combinar de manera fructífera con el templado simulado. Una forma es aplicar la característica de *oscilación estratégica* de la búsqueda tabú al programa de temperatura del templado simulado. Dicha oscilación ajusta el programa de temperatura al mover las temperaturas en un sentido y otro a través de los niveles donde se encontraron las mejores soluciones. Otra forma implica la aplicación de las estrategias de la lista de candidatos de la búsqueda tabú a la regla de selección del movimiento del templado simulado. La idea aquí es identificar múltiples vecinos para ver si se encuentra un movimiento con mejora antes de aplicar la regla aleatorizada para aceptar o rechazar el candidato actual a ser la próxima solución de prueba. Algunas veces, estos cambios producen mejoras significativas.

Como estas ideas para la aplicación de características de la búsqueda tabú al templado simulado lo sugieren, algunas veces un *algoritmo híbrido* que combina las ideas de diferentes metaheurísticas puede desempeñarse mejor que un algoritmo que se basa sólo en una sola metaheurística. Aunque en este capítulo se han presentado las tres metaheurísticas que se usan con mayor frecuencia por separado, en ocasiones los practicantes expertos escogen entre las ideas de éstas y otras metaheurísticas para diseñar sus métodos heurísticos.

Ejemplo de programación no lineal

Ahora reconsidere el ejemplo del problema de programación no lineal pequeño (con una sola variable) que se introdujo en la sección 13.1. El problema consiste en

$$\text{Maximizar} \quad f(x) = 12x^5 - 975x^4 + 28\,000x^3 - 345\,000x^2 + 1\,800\,000x,$$

sujeto a

$$0 \leq x \leq 31.$$

La gráfica de $f(x)$ en la figura 13.1 revela que existen óptimos locales en $x = 5$, $x = 20$ y $x = 31$, pero sólo $x = 20$ es un óptimo global.

El área de metaheurísticas del IOR Tutorial incluye una rutina para aplicar el algoritmo de templado simulado a problemas de programación no lineal pequeños de la forma,

$$\text{Maximizar} \quad f(x_1, \ldots, x_n)$$

sujeto a

$$L_j \leq x_j \leq U_j, \quad \text{para } j = 1, \ldots, n,$$

donde $n = 1$ o 2, y donde L_j y U_j son constantes ($0 \leq L_j < U_j \leq 63$) que representan los límites de x_j. (Resulta muy deseable tener límites estrechos para las variables individuales con el fin de aumentar la eficiencia de un algoritmo de templado simulado, lo cual también sucede en el caso de los algoritmos genéticos que se explican en la siguiente sección.) Cuando $n = 2$, también se puede incluir una o dos restricciones funcionales lineales sobre las variables $\mathbf{x} = (x_1, \ldots, x_n)$. En el ejemplo, se tiene

$$n = 1, \qquad L_1 = 0, \qquad U_1 = 31,$$

sin restricciones funcionales lineales.

Este procedimiento del IOR Tutorial diseña los detalles del algoritmo de templado simulado de los problemas de programación no lineal de este tipo como se describe a continuación:

1. **Solución de prueba inicial:** Se puede introducir cualquier solución factible, pero resulta útil introducir una que parezca ser buena. En ausencia de algunas pistas acerca de dónde podrían estar las buenas soluciones factibles, es razonable establecer cada variable x_j en el punto medio entre su límite inferior L_j y la cota superior U_j con el propósito de comenzar la búsqueda a la mitad de la región factible. (Por esta razón, $x = 15.5$ es una selección razonable de la solución de prueba inicial del ejemplo.)
2. **Estructura de vecindad:** Cualquier solución factible puede tomarse en cuenta para ser un vecino inmediato de la solución de prueba actual. Sin embargo, el método que se describe a continuación para seleccionar un vecino inmediato que se convierte en el candidato actual a ser la próxima solución de prueba, otorga preferencia a las soluciones factibles que están relativamente cerca de la solución de prueba actual. Aun así, permite la posibilidad de trasladarse a una parte diferente de la región factible para continuar la búsqueda.
3. **Selección aleatoria de un vecino inmediato:** Se establece

$$\sigma_j = \frac{U_j - L_j}{6}, \quad \text{para } j = 1, \ldots, n.$$

Después, dada la solución de prueba actual $(x_1, \ldots, x_n)$,

$$\text{se reescribe } x_j = x_j + N(0, \sigma_j), \quad \text{para } j = 1, \ldots, n,$$

donde $N(0, \sigma_j)$ es una observación aleatoria de una *distribución normal* con media cero y desviación estándar σ_j. Si este procedimiento no genera una solución factible, entonces se repite este procedimiento (se comienza de nuevo a partir de la solución de prueba actual) las veces que sea necesario para obtener una solución factible.

4. **Programa de temperatura:** Igual que en el caso de los problemas del agente viajero, se realizan cinco iteraciones a cada uno de los cinco valores de T (T_1, T_2, T_3, T_4, T_5), donde

$T_1 = 0.2Z_c$ cuando Z_c es el valor de la función objetivo de la solución de prueba inicial,
$T_2 = 0.5T_1$,
$T_3 = 0.5T_2$,
$T_4 = 0.5T_3$,
$T_5 = 0.5T_4$,

La razón para establecer $\sigma_j = (U_j - L_j)/6$ al seleccionar un vecino inmediato es que cuando la variable x_j está en el punto medio entre L_j y U_j, cualquier nuevo valor factible de la variable está dentro de tres desviaciones estándar del valor actual. Esto proporciona una probabilidad significativa de que el nuevo valor se mueva hacia uno de sus límites incluso cuando hay mucha mayor probabilidad de que el nuevo valor esté relativamente cerca del valor actual. Existen varios métodos para generar una observación aleatoria $N(0, \sigma_j)$ a partir de una distribución normal (como se explicará de manera breve en la sección 20.4). Por ejemplo, la función de Excel NORMINV(RAND(),0, σ_j) genera una observación aleatoria de este tipo. Para los problemas adicionales se presenta una forma directa de generar las observaciones aleatorias que se requieran. Obtenga un número aleatorio r y después use la tabla normal del apéndice 5 para encontrar el valor de $N(0, \sigma_j)$ tal que $P\{X \leq N(0, \sigma_j)\} = r$ cuando X es una variable aleatoria normal con media 0 y desviación estándar σ_j.

Para ilustrar la forma en que el algoritmo diseñado de esta forma se aplicaría al ejemplo, se comenzará con $x = 15.5$ como la solución de prueba inicial. Así,

$$Z_c = f(15.5) = 3\,741\,121 \quad \text{y} \quad T_1 = 0.2Z_c = 748\,224.$$

Como

$$\sigma = \frac{U - L}{6} = \frac{31 - 0}{6} = 5.167,$$

el siguiente paso es generar una observación aleatoria $N(0, 5.167)$ a partir de una distribución normal con media cero y esta desviación estándar. Para hacer esto, primero se obtiene un número aleatorio, que resulta ser 0.0735. Si se consulta la tabla normal en el apéndice 5, $P\{\text{normal estándar} \leq -1.45\} = 0.0735$, de manera que $N(0, 5.167) = -1.45(5.167) = -7.5$. El candidato actual a ser la siguiente solución de prueba se obtiene al reescribir x como

$$\begin{aligned} x &= 15.5 + N(0, 5.167) = 15.5 - 7.5 \\ &= 8, \end{aligned}$$

así que

$$Z_n = f(x) = 3\,055\,616.$$

Debido a que

$$\frac{Z_n - Z_c}{T} = \frac{3\,055\,616 - 3\,741\,121}{748\,224} = -0.916$$

la probabilidad de aceptar $x = 8$ como la próxima solución de prueba es

$$\text{Prob}\{\text{aceptación}\} = e^{-0.916} = 0.400.$$

Por tanto, $x = 8$ se aceptará sólo si el número aleatorio correspondiente entre 0 y 1 resulta ser menor que 0.400. En consecuencia, es bastante probable que $x = 8$ sea rechazada. (En algunas iteraciones posteriores cuando T es mucho más pequeña, $x = 8$ tendría casi la certeza de ser rechazada.) Esta perspectiva resulta afortunada puesto que la figura 13.1 revela que la búsqueda debería enfocarse en la parte de la región factible entre $x = 10$ y $x = 30$ con el fin de comenzar a escalar hacia la cumbre más alta.

En la tabla 13.6 se proporcionan los resultados que se obtuvieron al utilizar el IOR Tutorial para aplicar totalmente el algoritmo de templado simulado a este problema de programación no

TABLA 13.6 Aplicación del algoritmo de templado simulado del IOR Tutorial al ejemplo de programación no lineal

Iteración	*T*	Solución de prueba obtenida	*f*(*x*)
0		$x = 15.5$	3 741 121.0
1	748 224	$x = 17.557$	4 167 533.956
2	748 224	$x = 14.832$	3 590 466.203
3	748 224	$x = 17.681$	4 188 641.364
4	748 224	$x = 16.662$	3 995 966.078
5	748 224	$x = 18.444$	4 299 788.258
6	374 112	$x = 19.445$	4 386 985.033
7	374 112	$x = 21.437$	4 302 136.329
8	374 112	$x = 18.642$	4 322 687.873
9	374 112	$x = 22.432$	4 113 901.493
10	374 112	$x = 21.081$	4 345 233.403
11	187 056	$x = 20.383$	4 393 306.255
12	187 056	$x = 21.216$	4 330 358.125
13	187 056	$x = 21.354$	4 313 392.276
14	187 056	$x = 20.795$	4 370 624.01
15	187 056	$x = 18.895$	4 348 060.727
16	93 528	$x = 21.714$	4 259 787.734
17	93 528	$x = 19.463$	4 387 360.1
18	93 528	$x = 20.389$	4 393 076.988
19	93 528	$x = 19.83$	4 398 710.575
20	93 528	$x = 20.68$	4 378 591.085
21	46 764	$x = 20.031$	4 399 955.913 ← Máximo
22	46 764	$x = 20.184$	4 398 462.299
23	46 764	$x = 19.9$	4 399 551.462
24	46 764	$x = 19.677$	4 395 385.618
25	46 764	$x = 19.377$	4 383 048.039

lineal. Observe que las soluciones de prueba que se obtuvieron varían con bastante amplitud dentro de la región factible durante las primeras iteraciones, pero después comienzan a aproximarse a la cumbre más alta de una manera más coherente durante las últimas iteraciones cuando T se reduce a valores mucho más pequeños. Por tanto, de las 25 iteraciones, la mejor solución de prueba de $x = 20.031$ (en comparación con la solución óptima de $x = 20$) no se obtuvo hasta la iteración 21.

Una vez más, el lector puede encontrar interesante aplicar este software al mismo problema por sí mismo para ver qué se obtiene con las nuevas secuencias de números aleatorios y observaciones aleatorias a partir de distribuciones normales. (En el problema 13.3-6 se pide hacerlo varias veces.)

13.4 ALGORITMOS GENÉTICOS

Los algoritmos genéticos proporcionan un tercer tipo de metaheurística que es muy diferente de las primeras dos. Este tipo tiende a ser muy eficaz para explorar diversas partes de la región factible y evolucionar de manera gradual hacia las mejores soluciones factibles.

Después de introducir los conceptos básicos de este tipo de metaheurística, se aplicará un algoritmo genético básico al mismo ejemplo de programación no lineal que se consideró en la sección anterior con la restricción adicional de que la variable se limita a valores enteros. Después se aplicará este enfoque al mismo ejemplo del problema del agente viajero que se consideró en cada una de las secciones precedentes.

Conceptos básicos

De la misma forma que el templado simulado se basa en una analogía con un fenómeno natural (el proceso físico de templado), los algoritmos genéticos están muy influidos por otra forma de un fenómeno natural. En este caso, la analogía es con la *teoría biológica de la evolución* formulada por Charles Darwin a mediados del siglo XIX. Cada especie de plantas y animales muestra grandes

variaciones individuales. Darwin observó que aquellos individuos con variaciones que significan una ventaja de supervivencia a través de una mejoría en la adaptación al entorno tienen una posibilidad mayor de sobrevivir en la siguiente generación. Este fenómeno se conoce desde entonces como la *supervivencia del más apto*.

El moderno campo de la genética proporciona una mejor explicación de este proceso de evolución y de la *selección natural* involucrada en la supervivencia del más apto. En algunas especies que se reproducen por vía sexual, cada descendiente hereda algunos de los *cromosomas* de cada uno de los dos padres, donde los *genes* de los cromosomas determinan las características individuales del hijo. Un hijo que hereda las mejores características de los padres tiene una mayor posibilidad de sobrevivir en su adultez y se convierte en un padre que pasa algunas de estas características a la siguiente generación. La población tiende a mejorar de manera lenta a través del tiempo por medio de este proceso. Un segundo factor que contribuye a este proceso es una tasa de mutación aleatoria y de bajo nivel en el DNA de los cromosomas. En este contexto, de vez en cuando ocurre una *mutación* que cambia las características de un cromosoma que un hijo hereda de un padre. Aunque la mayoría de las mutaciones no tiene ningún efecto o son desventajosas, algunas proporcionan mejoras deseables. Los hijos con mutaciones deseables tienen una posibilidad un poco mayor de sobrevivir para así contribuir a la reserva futura de genes de la especie.

Estas ideas se transfieren hacia los problemas de optimización en una forma bastante natural. Las soluciones factibles de un problema específico corresponden a los miembros de una especie particular, donde la aptitud de cada miembro ahora se mide por el valor de la función objetivo. En lugar de procesar una sola solución de prueba a la vez (como sucede con las formas básicas de la búsqueda tabú y el templado simulado), ahora se trabaja con una *población* completa de soluciones de prueba.[1] En el caso de cada iteración (generación) de un algoritmo genético, la **población** actual consiste en el conjunto de soluciones de prueba que en la actualidad están bajo consideración. Estas soluciones de prueba se entienden como los miembros vivos de la especie. Algunos de los miembros más jóvenes de la población (en especial los miembros más aptos) sobreviven en la adultez y se convierten en **padres** (aparejados de manera aleatoria) que después tienen **hijos** (nuevas soluciones de prueba) que tienen algunas de las características (genes) de ambos padres. Como los miembros más aptos de la población tienen una mayor probabilidad de convertirse en padres que los otros, a medida que avanza, un algoritmo genético tiende a generar *poblaciones mejoradas* de soluciones de prueba. De vez en cuando ocurren **mutaciones**, de forma que los hijos también pueden adquirir características (algunas veces deseables) que no posee ninguno de los padres. Este fenómeno ayuda a los algoritmos genéticos a explorar una parte de la región factible, quizá mejor que la que se consideró antes. Por último, la supervivencia del más apto tiende a conducir al algoritmo genético hacia una solución de prueba (la mejor de todas las consideradas) que al menos es cercana a la óptima.

Aunque la analogía con el proceso de la evolución biológica define la esencia de cualquier algoritmo genético, no es necesario adherirse de manera rígida a esta analogía en cada detalle. Por ejemplo, algunos algoritmos genéticos (incluido el que se describe a continuación) permiten que la misma solución de prueba sea un padre en forma repetida a través de múltiples generaciones. De modo que la analogía debe ser sólo un punto de partida para definir los detalles del algoritmo que se ajuste mejor al problema bajo consideración.

A continuación se presenta un esquema bastante típico de un algoritmo genético que se empleará para los dos ejemplos.

Esquema de un algoritmo genético básico

Paso inicial. Comience con una población de soluciones de prueba factibles, quizá generándolas en forma aleatoria. Evalúe la *aptitud* (el valor de la función objetivo) de cada miembro de esta población actual.

Iteración. Use un proceso aleatorio que esté sesgado hacia los miembros más aptos de la población actual para seleccionar algunos de ellos (un número par) para convertirse en padres. Apareje

[1] Una de las estrategias de intensificación de la búsqueda tabú también mantiene una población de las mejores soluciones. La población se utiliza para crear rutas de vinculación entre sus miembros y para reiniciar la búsqueda a lo largo de estas rutas.

los padres de manera aleatoria para que de ellos nazcan dos hijos (nuevas soluciones de prueba *factibles*) cuyas características (genes) sean una mezcla aleatoria de las características de los padres, excepto por mutaciones ocasionales. (Cuando la mezcla aleatoria de características y algunas mutaciones generan una solución *no factible*, ésta se considera un *aborto*, y el proceso de intentar un nacimiento se repite hasta que nace un hijo que corresponde a una solución *factible*.) Conserve los hijos y una cantidad suficiente de los mejores miembros de la población actual para formar la nueva población del mismo tamaño de la próxima iteración. (Descarte los otros miembros de la población actual.) Evalúe la aptitud de cada nuevo miembro (hijo) de la población.

Regla de detención. Utilice alguna regla de detención, como un número fijo de iteraciones, una cantidad de tiempo de CPU fija o un número fijo de iteraciones consecutivas sin ninguna mejora a la mejor solución de prueba que haya encontrado hasta ese momento. Use la mejor solución de prueba que encontró en cualquier iteración como la solución final.

Antes de que este algoritmo se pueda implantar, es necesario contestar las siguientes preguntas:

1. ¿Cuál debe ser el tamaño de la población?
2. ¿Cómo deben seleccionarse los miembros de la población actual para convertirse en padres?
3. ¿Cuáles deben ser las características de los hijos derivadas de las características de los padres?
4. ¿Cómo deben insertarse las mutaciones en las características de los hijos?
5. ¿Cuál es la regla de detención que debe usarse?

Las respuestas a estas preguntas dependen en gran medida de la estructura del problema específico que se trata de resolver. En el área de metaheurísticas del IOR Tutorial se incluyen dos versiones del algoritmo. Una de ellas sirve para resolver problemas muy pequeños de programación no lineal entera como el ejemplo que se considera a continuación. La otra es para manejar problemas pequeños del agente viajero. Ambas versiones contestan algunas de las preguntas del mismo modo, como se describe en seguida.

1. **Tamaño de población:** Diez. (Este tamaño es razonable en los pequeños problemas para los que está diseñado este software, pero para problemas más grandes se usan poblaciones mucho más extensas.)
2. **Selección de padres:** De entre los cinco miembros más aptos de la población (de acuerdo con el valor de la función objetivo), seleccione cuatro de manera aleatoria para convertirse en padres. De entre los cinco miembros menos aptos, seleccione dos en forma aleatoria para volverse padres. Apareje los seis padres de manera aleatoria para formar tres parejas.
3. **Transferencia de características (genes) de padres a hijos:** En gran medida, este proceso depende del problema y por ende es diferente para las dos versiones del algoritmo del software, como se describe después para los dos ejemplos.
4. **Tasa de mutación:** La probabilidad de que una característica heredada de un hijo mute hacia una característica opuesta se establece en el software como 0.1. (En el caso de problemas más grandes es común utilizar tasas de mutación mucho más pequeñas.)
5. **Regla de detención:** El algoritmo se detiene después de cinco iteraciones sin ninguna mejora a la mejor solución de prueba que se haya encontrado hasta ese momento.

Ahora es posible aplicar el algoritmo a los dos ejemplos.

Versión entera del ejemplo de programación no lineal

Considere de nuevo el pequeño problema de programación no lineal que se introdujo en la sección 13.1 (vea la figura 13.1) y que después se resolvió con un algoritmo de templado simulado al final de la sección anterior. Ahora, a este ejemplo se le agrega la restricción de que la única variable x del problema debe tener un valor entero. Como el problema ya tiene la restricción de que $0 \leq x \leq 31$, lo cual significa que el problema tiene 32 soluciones factibles, $x = 0, 1, 2, \ldots, 31$. (La existencia de estos límites es muy importante para un algoritmo genético, puesto que reduce el espacio de búsqueda a la región relevante.) Por lo tanto, el ejemplo se ha convertido en un problema de programación no lineal *entera*.

Cuando se aplica un algoritmo genético, con frecuencia se utilizan *cadenas de dígitos binarios* para representar las soluciones del problema. Tal *codificación* de las soluciones es particularmente conveniente para los diferentes pasos de un algoritmo genético, incluido el proceso de nacimiento de los hijos. Esta codificación es fácil de hacer para este problema en particular, porque simplemente se puede escribir cada valor de x en base 2. Como el máximo valor factible de x es 31, sólo se requieren cinco dígitos binarios para escribir cualquier valor factible. Siempre se incluirán los cinco dígitos binarios aunque el primero o los primeros dígitos sean ceros. Así, por ejemplo,

$x = 3$ es 00011 en base 2,
$x = 10$ es 01010 en base 2,
$x = 25$ es 11001 en base 2.

Cada uno de los cinco dígitos binarios se conoce como uno de los **genes** de la solución, donde los dos valores posibles del dígito binario describen cuál de las dos posibles características se conserva en ese gen para ayudar a formar la composición genética global. Cuando los dos padres tienen la misma característica, éste se pasará a cada hijo (excepto cuando ocurre una mutación). Sin embargo, cuando los dos padres tienen características opuestas del mismo gen, la característica que heredará el hijo se vuelve aleatoria.

Por ejemplo, suponga que los dos padres son

P1: 00011 y
P2: 01010.

Como el primero, tercero y cuarto dígitos coinciden, en forma automática los hijos serán (salvo mutaciones)

C1: 0x01x y
C2: 0x01x,

donde x indica que este dígito particular todavía no se conoce. Para identificar los dígitos desconocidos se utilizan números aleatorios, donde una correspondencia natural es

0.0000-0.4999 corresponde a que el dígito sea 0.
0.5000-0.9999 corresponde a que el dígito sea 1.

Por ejemplo, suponga que los siguientes cuatro números aleatorios generados son 0.7265, 0.5190, 0.0402 y 0.3639, de manera que los dos dígitos desconocidos del primer hijo son 1 y los dos dígitos desconocidos del segundo hijo son 0. En este caso los hijos serán (salvo mutaciones)

C1: 01011 y
C2: 00010.

Este método particular para generar hijos a partir de los padres se conoce como *cruce uniforme*. Quizá es el más intuitivo de los diferentes métodos alternativos que se han propuesto.

Ahora es necesario considerar la posibilidad de mutaciones que pudieran afectar la conformación genética del hijo.

Como la probabilidad de una mutación en algún gen (que cambia el valor de un dígito binario al valor opuesto) se ha establecido en 0.1 para este algoritmo, se pueden utilizar de nuevo números aleatorios, donde

0.0000-0.0999 corresponde a una mutación,
0.1000-0.9999 corresponde a que no hay mutación.

Por ejemplo, suponga que en los siguientes 10 números aleatorios generados, sólo el octavo es menor a 0.1000. Esto indica que no ocurre mutación en el primer hijo, pero el tercer gen (dígito) del segundo hijo invierte su valor. Por tanto, la conclusión final es que los dos hijos son

C1: 01011 y
C2: 00110.

Si se regresa a la base 10, los dos padres corresponden a las soluciones $x = 3$ y $x = 10$, mientras que sus hijos serían (salvo mutaciones) $x = 11$ y $x = 2$. Sin embargo, debido a la mutación, los hijos se convierten en $x = 11$ y $x = 6$.

En este ejemplo particular, cualquier valor entero de x tal que $0 \leq x \leq 31$ (en base 10) es una solución factible, por lo que cada número de 5 dígitos en base 2 también es una solución factible. Por tanto, el proceso anterior de crear hijos nunca resulta en un *aborto* (una solución factible). Sin embargo, si el límite superior sobre x fuera, por ejemplo, $x \leq 25$, entonces ocurrirán abortos de manera ocasional. Siempre que sucede un aborto se descarta la solución y se repite el proceso completo de creación de un hijo hasta que se obtiene una solución factible.

Este ejemplo incluye sólo una variable. Para problemas de programación no lineal con múltiples variables, cada miembro de la población usaría de nuevo la base dos para mostrar el valor de cada variable. Entonces, el proceso anterior para generar hijos a partir de los padres se realizaría de la misma forma para una variable a la vez.

En la tabla 13.7 se muestra la aplicación completa del algoritmo a este ejemplo a través del paso inicial [parte *a*) de la tabla] y de la iteración 1 [parte *b*) de la tabla]. En el paso de inicio, cada uno de los miembros de la población inicial se generó mediante cinco números aleatorios y la utilización de la correspondencia entre un número aleatorio y un dígito binario, que se dio con anterioridad para obtener los cinco dígitos binarios a la vez. El valor correspondiente de x en base 10 se inserta en la función objetivo que se dio al inicio de la sección 13.1 para evaluar la aptitud de ese miembro de la población.

Los cinco números de la población inicial que tienen el grado más alto de aptitud (en orden) son los miembros 10, 8, 4, 1 y 7. Para elegir en forma aleatoria cuatro de estos miembros para convertirse en padres, se usa un número aleatorio para seleccionar un miembro que será rechazado, donde 0.0000-0.1999 significa la expulsión del primer miembro de la lista (miembro 10), 0.2000-0.3999 corresponde a expulsar el segundo miembro, y así sucesivamente. En este caso, el número aleatorio fue 0.9665 (miembro 7), por lo que el quinto miembro de la lista no se convierte en padre.

De entre los cinco miembros menos aptos de la población inicial (miembros 2, 1, 6, 5 y 9), ahora se usan números aleatorios para seleccionar cuáles dos de estos miembros se convertirán en padres. En este caso los números aleatorios fueron 0.5634 y 0.1270. Con referencia al primer número aleatorio, 0.0000-0.1999 corresponde seleccionar el primer número de la lista (miembro 2), 0.2000-0.3999 corresponde a seleccionar el segundo miembro, y así sucesivamente, por lo que

TABLA 13.7 Aplicación del algoritmo genético al ejemplo de programación no lineal entera durante *a*) el paso inicial y *b*) la iteración 1

	Miembro	Población inicial	Valor de *x*	Aptitud
a)	1	0 1 1 1 1	15	3 628 125
	2	0 0 1 0 0	4	3 234 688
	3	0 1 0 0 0	8	3 055 616
	4	1 0 1 1 1	23	3 962 091
	5	0 1 0 1 0	10	2 950 000
	6	0 1 0 0 1	9	2 978 613
	7	0 0 1 0 1	5	3 303 125
	8	1 0 0 1 0	18	4 239 216
	9	1 1 1 1 0	30	1 350 000
	10	1 0 1 0 1	21	4 353 187

	Miembro	Padres	Hijo	Valor de *x*	Aptitud
b)	10	1 0 1 0 1	0 0 1 0 1	5	3 303 125
	2	0 0 1 0 0	1 0 0 0 1	17	4 064 259
	8	1 0 0 1 0	1 0 0 1 1	19	4 357 164
	4	1 0 1 1 1	1 0 1 0 0	20	4 400 000
	1	0 1 1 1 1	0 1 0 1 1	11	2 980 637
	6	0 1 0 0 1	0 1 1 1 1	15	3 628 125

en este caso el miembro seleccionado es el tercero (miembro 6). Como ahora restan sólo cuatro números (2, 1, 5 y 9) para seleccionar el último padre, los intervalos correspondientes al segundo número aleatorio son 0.0000-0.2499, 0.2500-0.4999, 0.5000-0.7499 y 0.7500-0.9999. Como 0.1270 cae en el primero de estos intervalos se selecciona el primer miembro de la lista (miembro 2) para ser un padre.

El siguiente paso es aparejar los seis padres, esto es, los miembros 10, 8, 4, 1, 6 y 2. Se comenzará por usar un número aleatorio para determinar la pareja del primer miembro de la lista (miembro 10). El número aleatorio 0.8204 indicó que 10 debe estar aparejado con el quinto de los otros cinco padres de la lista (miembro 2). Para aparejar el siguiente miembro de la lista (miembro 8), el número aleatorio fue 0.0198, que se encuentra en el intervalo 0.0000-0.3333, por lo que el primero de los tres padres restantes (miembro 4) se elige para ser el compañero del miembro 8. Con esto restan dos padres (miembros 1 y 6) que se convierten en la última pareja.

En la parte *b*) de la tabla 13.7 se muestran los hijos reproducidos por estos padres por medio del proceso que se ilustró antes en esta subsección. Observe las mutaciones que sufrió el tercer gen del segundo hijo y el cuarto gen del cuarto hijo. Asimismo, el sexto hijo tiene un grado de aptitud relativamente alto. En realidad, para cada pareja, ambos hijos resultaron ser más aptos que uno de los padres. Esto no sucede siempre, pero es bastante común. En el caso de la segunda pareja de padres, los dos hijos resultaron ser más aptos que ambos padres. De manera fortuita, los dos hijos ($x = 19$ y $x = 20$) en realidad son superiores a *cualquiera* de los miembros de la población anterior que se presentó en la parte *a*) de la tabla. Para formar la nueva población de la próxima iteración, los seis hijos se conservan junto con los cuatro miembros más aptos de la población anterior (miembros 10, 8, 4 y 1).

Las iteraciones siguientes se realizan de un modo similar. Como ya se sabe con base en la explicación de la sección 13.1 (vea la figura 13.1) que $x = 20$ (la mejor solución de prueba generada en la iteración 1) en realidad es la solución óptima de este ejemplo, las iteraciones subsecuentes no proporcionarían ninguna mejora. Por tanto, la regla de detención terminaría el algoritmo después de cinco iteraciones más y se obtendría $x = 20$ como la solución final.

En el IOR Tutorial se incluye una rutina para aplicar este mismo algoritmo genético a otros problemas muy pequeños de programación no lineal entera. (La forma y el tamaño de las restricciones son los mismos que los que se especificaron en la sección 13.3 para el caso de problemas de programación no lineal.)

El lector podría encontrar interesante la aplicación de esta rutina del IOR Tutorial a este mismo ejemplo. Debido a la aleatoriedad inherente del algoritmo, se obtienen resultados intermedios diferentes cada vez que éste se aplica. (En el problema 13.4-3 se pide aplicar el algoritmo a este ejemplo varias veces.)

Aunque éste fue un ejemplo discreto, los algoritmos genéticos también se pueden aplicar a problemas continuos, como un problema de programación no lineal sin la restricción de valores enteros. En este caso, el valor de una variable continua se representaría (o tendría una aproximación cercana) por medio de un número decimal en base 2. Por ejemplo, $x = 23\frac{5}{8}$ es 10111.10100 en base 2, y $x = 23.66$ es muy cercano a 10111.10101 en base 2. Todos los dígitos binarios de ambos lados del punto decimal se pueden tratar como se hizo antes para aparejar padres que producen hijos, etcétera.

Ejemplo del problema del agente viajero

En las secciones 13.2 y 13.3 se ilustró cómo se aplican los algoritmos de búsqueda tabú y de templado simulado al problema del agente viajero particular que se introdujo en la sección 13.1 (vea la figura 13.4). Ahora se verá cómo puede aplicarse el algoritmo genético a este mismo ejemplo.

En este caso, en lugar de utilizar dígitos binarios se continuará con la representación de cada solución (viaje) en la forma natural como una secuencia de ciudades visitadas. Por ejemplo, la primera solución que se considera en la sección 13.1 es el viaje por las ciudades en el siguiente orden: 1-2-3-4-5-6-7-1, donde la ciudad 1 es la población de residencia en la que el viaje debe iniciarse y terminar. Sin embargo, debe puntualizarse que los algoritmos genéticos para manejar los problemas del agente viajero con frecuencia utilizan otros métodos para *codificar* soluciones. En general, los métodos inteligentes para representar soluciones (a menudo con la utilización de cadenas de dígitos binarios) pueden hacer más fácil la generación de hijos, la creación de mutaciones, la conservación de factibilidad, etc., de un modo natural. El desarrollo de un *esquema de*

codificación apropiado es una parte clave de la elaboración de un algoritmo genético eficaz para cualquier aplicación.

Una complicación con este ejemplo particular es que, en cierto sentido, es muy fácil. Debido al número bastante limitado de ligaduras entre los pares de ciudades de la figura 13.4, este problema apenas alcanza 10 soluciones factibles distintas si se descartan los viajes que son sólo un viaje considerado antes pero en la dirección inversa. Por tanto, no es posible tener una población inicial con 10 distintas soluciones de prueba tales que los seis padres resultantes reproduzcan hijos diferentes que también sean distintos de los miembros de la población inicial (incluidos los padres).

Por fortuna, un algoritmo genético puede aun operar en forma razonable cuando existe una tasa modesta de duplicación de las soluciones de prueba en una población o en dos poblaciones consecutivas. Por ejemplo, aun cuando los dos padres de una pareja sean idénticos, es posible que los hijos sean diferentes a los padres debido a las mutaciones.

El algoritmo genético para manejar problemas del agente viajero del IOR Tutorial no hace nada para evitar la duplicación de las soluciones de prueba consideradas. Cada una de las 10 soluciones de prueba de la población inicial se genera de la manera siguiente. A partir de la ciudad de residencia se usan números aleatorios para seleccionar la siguiente ciudad de entre aquellas que tienen una ligadura con la ciudad inicial (2, 3 y 7 en la figura 13.4). Después se usan números aleatorios para seleccionar la tercera ciudad de entre las ciudades restantes que tienen una ligadura con la segunda ciudad. Este proceso se continúa hasta que todas las ciudades se incluyen una vez en el viaje (más un regreso a la ciudad de residencia a partir de la última ciudad) o se llega a un final anticipado porque no existen ligaduras de la ciudad actual con ninguna de las ciudades que aún no se han visitado. En el último caso se reinicia el proceso completo para generar una solución de prueba, desde el principio y con números aleatorios nuevos.

También se usan números aleatorios para reproducir hijos a partir de una pareja de padres. Para ilustrar este proceso, considere el siguiente par de padres.

P1: 1-2-3-4-5-6-7-1
P2: 1-2-4-6-5-7-3-1

Conforme se describa el proceso de generación de hijos a partir de estos padres, se presentará un resumen de estos resultados en la tabla 13.8 para ayudarle a seguir el avance.

Si por el momento no se toma en cuenta la posibilidad de mutaciones, a continuación se describe la idea fundamental de cómo generar un hijo.

> **Enlaces heredados:** Los genes corresponden a los enlaces en un viaje. Entonces cada uno de los genes (ligaduras) heredado por un hijo debe provenir de uno de los padres (o de ambos). (La excepción descrita en el próximo párrafo es que un padre también puede pasar a su hijo un subviaje inverso.) Debido a que son heredados, estos enlaces se seleccionan aleatoriamente uno a la vez hasta que se genera un viaje completo (el hijo).

Para comenzar este proceso con los padres mencionados antes, como un viaje debe comenzar en la ciudad 1, la ligadura inicial de un hijo debe provenir de una de las ligaduras de los padres que

TABLA 13.8 Ejemplo del proceso de generación de un hijo para el caso del agente viajero

Padre P1: 1-2-3-4-5-6-7-1
Padre P2: 1-2-4-6-5-7-3-1

Ligadura	Opciones	Selección aleatoria	Viaje
1	1-2, 1-7, 1-2, 1-3	1-2	1-2
2	2-3, 2-4	2-4	1-2-4
3	4-3, 4-5, 4-6	4-3	1-2-4-3
4	3-5*, 3-7	3-5*	1-2-4-3-5
5	5-6, 5-6, 5-7	5-6	1-2-4-3-5-6
6	6-7	6-7	1-2-4-3-5-6-7
7	7-1	7-1	1-2-4-3-5-6-7-1

*Ligadura que termina un subviaje inverso.

conectan la ciudad 1 con otra ciudad. En el caso del padre P1, éstas son las ligaduras 1-2 y 1-7. (La ligadura 1-7 califica puesto que es equivalente tomar el viaje en cualquier dirección.) En el del padre P2, las ligaduras correspondientes son 1-2 (de nuevo) y 1-3. El hecho de que ambos padres tengan la ligadura 1-2 duplica la probabilidad de que ésta sea heredada por un hijo. Por lo tanto, cuando se usa un número aleatorio para determinar cuál ligadura heredará el hijo, el intervalo 0.0000-0.4999 (o cualquier intervalo de este tamaño) corresponde a heredar la ligadura 1-2 mientras que los intervalos 0.50000-0.7499 y 0.7500-0.9999 corresponderían a la elección de la ligadura 1-7 y la ligadura 1-3, respectivamente. Suponga que se elige 1-2 como se muestra en el primer renglón de la tabla 13.8. Después de 1-2, un padre utiliza en seguida la ligadura 2-3 mientras que el otro usa 2-4. Por tanto, al generar el hijo, se debe hacer una selección aleatoria entre estas dos opciones. Suponga que se selecciona 2-4 (vea el segundo renglón de la tabla 13.8). Ahora existen tres opciones para la ligadura que sigue a 1-2-4 porque el primer padre usa dos ligaduras (4-3 y 4-5) para conectar a la ciudad 4 en su viaje y el segundo padre usa la ligadura 4-6 (la ligadura 4-2 se pasa por alto porque la ciudad 2 ya está en el viaje del hijo). Al seleccionar de manera aleatoria una de estas opciones, suponga que se elige 4-3 para formar 1-2-4-3 como el inicio del viaje del hijo hasta ahora, como se muestra en el tercer renglón de la tabla 13.8.

Ahora se llega a una característica adicional de este proceso para generar el viaje de un hijo, a saber, el uso de un *subviaje inverso* del padre.

> **Subviaje inverso heredado:** Otra posibilidad de un enlace heredado por un hijo es un enlace que es necesario para completar un subviaje inverso al del viaje que el hijo está haciendo en una parte de un viaje del padre.

Para ilustrar cómo puede surgir esta posibilidad, observe que la siguiente ciudad después de 1-2-4-3 debe ser una de las ciudades que no se ha visitado (ciudad 5, 6 o 7). Pero el primer padre no tiene ninguna ligadura de la ciudad 3 a alguna de estas otras ciudades. La razón es que el hijo utiliza un subviaje inverso (al invertir 3-4) del viaje de este padre, 1-2-3-4-5-6-7-1. Para completar este subviaje inverso se requiere agregar la ligadura 3-5, con lo cual ésta se convierte en una de las opciones para la siguiente ligadura del viaje del hijo. La otra opción es la ligadura 3-7 que proporciona el segundo padre (la ligadura 3-1 no se considera una opción porque la ciudad 1 debe estar hasta el final del viaje). Una de estas dos opciones se selecciona de manera aleatoria. Suponga que la elección es la ligadura 3-5, lo cual proporciona el viaje del hijo hasta el momento, 1-2-4-3-5, como se muestra en el cuarto renglón de la tabla 13.8.

Para continuar este viaje, las opciones de la siguiente ligadura son 5-6 (que proporcionan los dos padres) y 5-7 (que proporciona el segundo padre). Suponga que la elección aleatoria entre 5-6, 5-6 y 5-7 es 5-6, por lo que el viaje hasta ahora es 1-2-4-3-5-6. (Vea la quinta hilera de la tabla 13.8.) Como la única ciudad que aún no se visita es la 7, se agrega de manera automática la ligadura 6-7 seguida por la ligadura 7-1 para regresar a la ciudad de residencia. De esta forma, el viaje completo del hijo es

C1: 1-2-4-3-5-6-7-1

En la figura 13.5 de la sección 13.1 se observa qué tanto se parece este hijo al primer padre, puesto que la única diferencia es el subviaje inverso que se obtuvo al invertir 3-4 en el viaje del padre.

Si se hubiera elegido la ligadura 5-7 para agregarse a 1-2-4-3-5, el viaje se hubiera completado de manera automática como 1-2-4-3-5-7-6-1. Sin embargo, no existe una ligadura 6-1 (vea la figura 13.4), por lo que se llega a un final anticipado en la ciudad 6. Cuando esto sucede, ocurre un *aborto* y es necesario reiniciar todo el proceso desde el principio y con números aleatorios nuevos hasta que se obtenga un hijo con un viaje completo. Después se repite este proceso para obtener el segundo hijo.

Ahora es necesario agregar una característica más —la posibilidad de mutaciones— para completar la descripción del proceso de generación de hijos.

> **Mutaciones de enlaces heredados:** Siempre que un enlace particular sea heredado a partir del padre de un hijo, existe una reducida probabilidad de que ocurra una mutación que rechace ese enlace y, en lugar de ello, seleccione aleatoriamente uno de los demás enlaces de la ciudad actual a otra ciudad que no esté en el viaje, sin considerar si dicho enlace es usado por cualquiera de los padres.

El algoritmo genético para los problemas del agente viajero implantado en su tutorial IOR utiliza una probabilidad de 0.1 de que ocurra una mutación cada vez que sea necesario seleccionar el enlace siguiente en el viaje del hijo. Así, siempre que el número aleatorio correspondiente sea menor a 0.1000, la elección de la ligadura que se realizó de la forma normal que se describió con anterioridad se rechaza (si existe alguna otra elección posible). En lugar de ello, todas las otras ligaduras que parten de la ciudad actual hacia una ciudad que aún no está en el viaje (incluso ligaduras no proporcionadas por los padres) se identifican, y una de ellas se selecciona en forma aleatoria para ser la siguiente ligadura del viaje. Por ejemplo, suponga que ocurre una mutación al generar la primera ligadura del hijo. A pesar de que se había elegido 1-2 de manera aleatoria como la primera ligadura, ahora ésta se debería rechazar debido a la mutación. En razón de que la ciudad 1 también tiene ligas con las ciudades 3 y 7 (vea la figura 13.4), cualquiera de las ligaduras 1-3 o 1-7 se selecciona en forma aleatoria para ser el primer viaje. (Como los padres terminan sus viajes con el uso de una u otra de estas ligaduras, este caso se puede entender como el inicio del viaje del hijo invirtiendo la dirección de los viajes de uno de los padres.)

Ahora es posible escribir un esquema del procedimiento general para generar un hijo a partir de una pareja de padres.

Procedimiento para generar un hijo

1. **Paso inicial:** Para comenzar, designe la ciudad de residencia como la *ciudad actual*.
2. **Opciones de la siguiente ligadura:** Identifique todas las ligaduras que van de la ciudad actual a otra ciudad, que todavía no están en el viaje del hijo y que sean utilizadas por los padres en cualquier dirección. Además, agregue cualquier ligadura que se necesite para completar un subviaje inverso que realiza el hijo en una parte del viaje del padre.
3. **Selección de la siguiente ligadura:** Use un número aleatorio para seleccionar de manera aleatoria una de las opciones identificadas en el paso 2.
4. **Verificación de una mutación:** Si el siguiente número aleatorio es menor que 0.1000, ocurre una mutación y la ligadura seleccionada en el paso 3 se rechaza (a menos que no exista otra ligadura desde la ciudad actual hasta otra ciudad que no se encuentre en el viaje). Si la ligadura se rechaza, identifique todas las otras ligaduras que parten de la ciudad actual y van a otra ciudad que no esté en el viaje del hijo (lo cual incluye las ligaduras no utilizadas por ningún padre). Use un número aleatorio para seleccionar de manera aleatoria una de estas otras ligaduras.
5. **Continuación:** Agregue la ligadura que seleccionó en el paso 3 (si no ocurre mutación) o en el paso 4 (si hay mutación) al final del viaje incompleto actual del hijo y reasigne la ciudad al final de esta ligadura como la *ciudad actual*. Si aún queda más de una ciudad no incluida en el viaje (además del regreso a la ciudad de residencia), regrese a los pasos 2-4 para seleccionar la siguiente ligadura. De otra forma, vaya al paso 6.
6. **Terminación:** Cuando sólo reste una ciudad sin incluir en el viaje del hijo, agregue la ligadura desde la ciudad actual hasta esta ciudad restante. Después agregue la ligadura que va de la última ciudad a la ciudad de residencia para completar el viaje del hijo. Sin embargo, si la ligadura necesaria no existe ocurre un aborto y el procedimiento debe reiniciarse de nuevo desde el paso 1.

Este procedimiento se aplica para cada par de padres para obtener cada uno de los dos hijos.

El algoritmo genético para problemas del agente viajero del IOR Tutorial incorpora este procedimiento para generar hijos como parte de todos los algoritmos descritos al comienzo de esta sección. En la tabla 13.9 se muestran los resultados de la aplicación de este algoritmo al ejemplo durante el paso inicial y la primera iteración de todo el algoritmo. Debido a la aleatoriedad que se incorpora al algoritmo, sus resultados intermedios (y quizá también la mejor solución final) variarán cada vez que se corra el algoritmo hasta su culminación. (Para explorar más a fondo esta cuestión, en el problema 13.4-7 se pide utilizar el IOR Tutorial para aplicar el algoritmo completo a este ejemplo varias veces.)

El hecho de que el ejemplo tenga sólo un número relativamente pequeño de soluciones factibles distintas se refleja en los resultados que se muestran en la tabla 13.9. Los miembros 1, 4, 6 y 10 son idénticos, al igual que los miembros 2, 7 y 9 (excepto porque el miembro 2 realiza su viaje en la dirección inversa). Por lo tanto, la generación aleatoria de 10 miembros de la población inicial

TABLA 13.9 Aplicación del algoritmo genético del IOR Tutorial al ejemplo del problema del agente viajero durante *a*) el paso inicial y *b*) la iteración 1

	Miembro	Población inicial	Distancia
a)	1	1-2-4-6-5-3-7-1	64
	2	1-2-3-5-4-6-7-1	65
	3	1-7-5-6-4-2-3-1	65
	4	1-2-4-6-5-3-7-1	64
	5	1-3-7-5-6-4-2-1	66
	6	1-2-4-6-5-3-7-1	64
	7	1-7-6-4-5-3-2-1	65
	8	1-3-7-6-5-4-2-1	69
	9	1-7-6-4-5-3-2-1	65
	10	1-2-4-6-5-3-7-1	64

	Miembro	Padres	Hijo	Miembro	Distancia
b)	1	1-2-4-6-5-3-7-1	1-2-4-5-6-7-3-1	11	69
	7	1-7-6-4-5-3-2-1	1-2-4-6-5-3-7-1	12	64
	2	1-2-3-5-4-6-7-1	1-2-4-5-6-7-3-1	13	69
	6	1-2-4-6-5-3-7-1	1-7-6-4-5-3-2-1	14	65
	4	1-2-4-6-5-3-7-1	1-2-4-6-5-3-7-1	15	64
	5	1-3-7-5-6-4-2-1	1-3-7-5-6-4-2-1	16	66

resultó en sólo cinco distintas soluciones factibles. En forma similar, cuatro de los seis hijos generados (miembros 12, 14, 15 y 16) son idénticos a uno de sus padres (excepto porque el miembro 14 realiza el viaje en la dirección opuesta a su primer padre). Dos de los hijos (miembros 12 y 15) tienen una mejor aptitud (distancia más corta) que uno de sus padres, pero ninguno superó a ambos padres. Ninguno de estos hijos proporcionó una solución óptima (que tiene una distancia de 63). Esto ilustra el hecho de que, para algunos problemas, un algoritmo genético puede requerir muchas generaciones (iteraciones) antes de que el fenómeno de la supervivencia del más apto resulte en la creación de poblaciones significativamente superiores.

En la sección de Worked Examples del sitio en internet de este libro, se proporciona **otro ejemplo** de la aplicación de este algoritmo genético a un problema del agente viajero. Este problema tiene un número más grande de soluciones factibles que el ejemplo anterior, por lo cual hay una gran diversidad en la población inicial, los padres resultantes y sus hijos.

13.5 CONCLUSIONES

Algunos problemas de optimización (incluidos diferentes problemas de optimización combinatoria) son tan complejos que resulta imposible resolverlos y encontrar una solución óptima con el tipo de algoritmos exactos que se presentaron en los capítulos anteriores. En tales casos, los métodos heurísticos se usan por lo general para buscar una buena solución factible (no necesariamente la óptima). Existen varias metaheurísticas que proporcionan una estructura general y directrices estratégicas para diseñar un método heurístico específico que se ajuste a un problema en particular. Una característica clave de estos procedimientos metaheurísticos es su capacidad de escapar de un óptimo local y realizar una búsqueda vigorosa en la región factible.

Este capítulo introdujo los tres tipos de metaheurísticas más prominentes. La *búsqueda tabú* se mueve desde la solución de prueba actual hasta la mejor solución de prueba vecina en cada iteración, procedimiento muy parecido a uno de mejora local, con la salvedad de que permite un movimiento sin mejora cuando no existen movimientos que mejoren la solución de prueba actual. También incorpora una memoria a corto plazo de la búsqueda pasada para incentivar el movimiento hacia nuevas áreas de la región factible en lugar de regresar a soluciones consideradas con anterio-

ridad. Además, puede utilizar estrategias de intensificación y diversificación basadas en la memoria a largo plazo para enfocar la búsqueda en continuaciones promisorias. El *templado simulado* también se desplaza desde la solución de prueba actual hasta una solución de prueba vecina en cada iteración mientras permite de manera ocasional movimientos sin mejora. Sin embargo, seleccione la solución de prueba vecina de manera aleatoria y después utilice la analogía con un proceso de templado físico para determinar si este vecino debe rechazarse como la próxima solución de prueba si no es tan bueno como la solución de prueba actual. El tercer tipo de metaheurística, los *algoritmos genéticos*, trabaja con una población completa de soluciones de prueba en cada iteración. Después utiliza la analogía con la teoría biológica de la evolución, sobre todo el concepto de la supervivencia del más apto, para descartar algunas de las soluciones de prueba (en especial las más pobres) y reemplazarlas con otras nuevas. Este proceso de reemplazo tiene pares de miembros sobrevivientes que transfieren algunas de sus características a pares de nuevos miembros como si fueran padres que reproducen hijos.

Con la intención de ser concretos, se describió el algoritmo básico de cada metaheurística y después se adaptó a dos tipos específicos de problemas (incluido el problema del agente viajero), mediante el uso de ejemplos simples. Sin embargo, los investigadores también han desarrollado muchas variaciones de cada algoritmo que son utilizadas por los practicantes para adaptarlas a las características de los problemas complejos a los que se enfrentan. Por ejemplo, se han propuesto docenas de variaciones del algoritmo genético básico para problemas del agente viajero (entre ellas diferentes procedimientos para generar hijos), y la investigación continúa para determinar cuál es la más eficaz. (Algunos de los mejores métodos para problemas del agente viajero usan estrategias especiales de "k-opt" y "cadena de expulsión" que se diseñan de manera cuidadosa para tomar ventaja de la estructura del problema.) Por lo tanto, las lecciones importantes de este capítulo son los conceptos básicos y la intuición incorporada en cada metaheurística más que los detalles de los algoritmos particulares que se presentaron.

Existen algunos otros tipos importantes de metaheurística además de los tres que se presentaron en este capítulo. Dentro de éstos, por ejemplo, se encuentran la optimización de la colonia de hormigas, la búsqueda esparcida y las redes neuronales artificiales. (Estos nombres sugestivos proporcionan una clave de la idea básica que está detrás de esta metaheurística.) La referencia seleccionada 4 ofrece una cobertura a fondo de esta metaheurística y de las tres que se presentaron aquí. (Michel Gendreau y Jean-Yves Potvin están preparando una segunda edición con el fin de actualizar esta importante referencia.)

En realidad, algunos algoritmos heurísticos son un híbrido de diferentes tipos de metaheurísticas que tratan de combinar sus mejores características. Por ejemplo, la búsqueda tabú a corto plazo (sin componente de diversificación) es muy eficiente para buscar óptimos locales pero no lo es tanto para explorar las diferentes áreas de la región factible para encontrar la parte que contiene el óptimo global, mientras que un algoritmo genético tiene las características opuestas. Por lo tanto, algunas veces se puede obtener un algoritmo mejorado si se comienza con un algoritmo genético para tratar de encontrar la cumbre más alta (cuando el objetivo es maximizar) y después se cambia a una búsqueda tabú básica para, por último, escalar hasta dicha cumbre. También existen otras metaheurísticas que son menos prominentes que las tres que se presentaron aquí, cuyas ideas básicas también podrían incorporarse en un algoritmo heurístico. La clave para diseñar un algoritmo heurístico eficaz es incorporar aquella idea que funcione mejor para el problema bajo consideración en lugar de adherirse en forma rígida a la filosofía de una metaheurística particular.

REFERENCIAS SELECCIONADAS

1. Coello, C., D. A. Van Veldhuizen y G. B. Lamont: *Evolutionary Algorithms for Solving Multi-Objective Problems*, Kluwer Academic Publishers (Springer en la actualidad), Boston, 2002.
2. Gen, M. y R. Cheng: *Genetic Algorithms and Engineering Optimization*, Wiley, Nueva York, 2000.
3. Glover, F.: "Tabu Search: A Tutorial", en *Interfaces*, **20**(4): 74-94, julio-agosto de 1990.
4. Glover, F. y G. Kochenberger (eds.): *Handbook of Metaheuristics*, Kluwer Academic Publishers (Springer en la actualidad), Boston, MA, 2003. (Esta referencia proporciona una cobertura actualizada de todas las metaheurísticas que se consideraron en este capítulo, así como otras metaheurísticas.)
5. Glover, F. y M. Laguna: *Tabu Search*, Kluwer Academic Publishers (Springer en la actualidad), Boston, MA, 1997.

6. Gutin, G. y A. Punnen (eds.): *The Traveling Salesman Problem and Its Variations*, Kluwer Academic Publishers (Springer en la actualidad), Boston, MA, 2002.
7. Haupt, R. L. y S. E. Haupt: *Practical Genetic Algorithms*, Wiley, Nueva York, 1998.
8. Jones, D. F., S. K. Mirrazavi y M.Tamiz: "Multiobjective Metaheuristics: An Overview of the Current State of the Art", en *European Journal of Operational Research*, **137:** 1-9, 2002.
9. Laguna M. y R. Marti: *Scatter Search: Methodology and Implementations* en C, Kluwer Academic Publishers (Springer en la actualidad), Boston, 2003.
10. Michalewicz, Z. y D. B. Fogel: *How To Solve It: Modern Heuristics*, Springer, Berlín, 2002.
11. Mitchell, M.: *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, MA, 1998.
12. Molina, J., M. Laguna, R. Marti y R. Caballero: "SSPMO: A Scatter Tabu Search Procedure for Non-Linear Multiobjective Optimization", en *INFORMS Journal on Computing*, **19**(1): 91-100, Winter, 2007.
13. Reeves, C. R.: "Genetic Algorithms for the Operations Researcher", en *INFORMS Journal on Computing*, **9:** 231-250, 1997. (También vea en pp. 251-265 comentarios sobre este artículo.)
14. Sarker, R., M. Mohammadian y X. Yao (eds.): *Evolutionary Optimization*, Kluwer Academic Publishers (Springer en la actualidad), Boston, MA, 2002.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN EL SITIO DE INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos para el capítulo 13

Rutinas automáticas en el IOR Tutorial:

Algoritmo de búsqueda tabú para problemas del agente viajero (Tabu Search Algorithm for Traveling Salesman Problems)
Algoritmo de templado simulado para problemas del agente viajero (Simulated Annealing Algorithm for Traveling Salesman Problems)
Algoritmo de templado simulado para problemas de programación no lineal (Simulated Annealing Algorithm for Nonlinear Programming Problems)
Algoritmo genético para problemas de programación no lineal entera (Genetic Algorithm for Integer Nonlinear Programming Problems)
Algoritmo genético para problemas del agente viajero (Genetic Algorithm for Traveling Salesman Problems)

Glosario del capítulo 13

Vea en el apéndice 1 la documentación del software.

PROBLEMAS

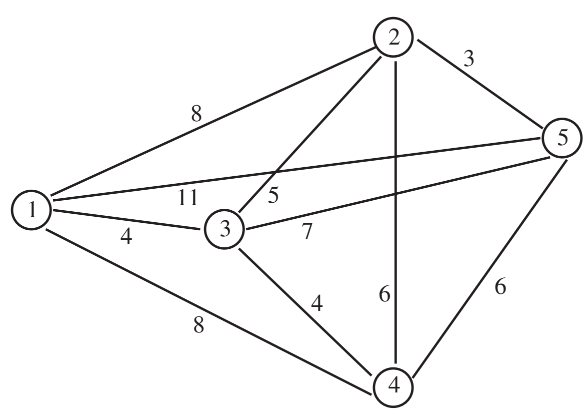
La letra A a la izquierda de algunos de los problemas (o sus incisos) tiene el siguiente significado.

A: Se puede utilizar la rutina automática del IOR Tutorial. La impresión registra su trabajo en cada iteración.

Un asterisco después del número del problema indica que al final del libro se da al menos una respuesta parcial.

Instrucciones para obtener números aleatorios
Para cada problema o sus incisos donde se necesiten números aleatorios, obténgalos de los dígitos aleatorios consecutivos de la tabla 20.3 de la sección 20.3 de la manera siguiente. Inicie en el renglón superior de la tabla y forme números aleatorios *de cinco dígitos* colocando un punto decimal enfrente de cada grupo de cinco dígitos aleatorios (0.09656, 0.96657, etc.) en el orden en que se requiera los números aleatorios. Siempre reinicie desde el renglón superior para cada nuevo problema o inciso.

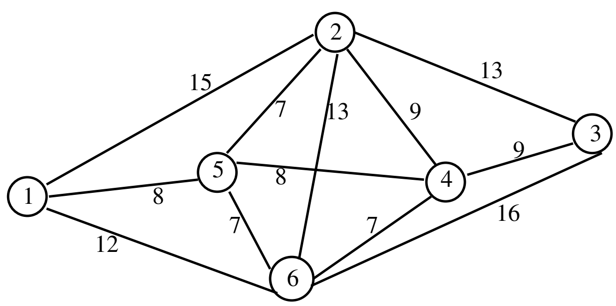
13.1-1. Considere el problema del agente viajero que se muestra a continuación, donde la ciudad 1 es la ciudad de residencia.



a) Enumere todos los viajes posibles; excluya únicamente aquellos que sólo inviertan el orden de los viajes de la lista que se presentó con anterioridad. Calcule la distancia de cada uno de estos viajes y con esto identifique el viaje óptimo.
b) Comience con 1-2-3-4-5-1 como la solución de prueba inicial para aplicar el algoritmo de subviaje inverso a este problema.
c) Aplique el algoritmo de subviaje inverso a este problema; para ello comience con 1-2-4-3-5-1 como la solución de prueba inicial.
d) Aplique el algoritmo de subviaje inverso a este problema; para ello comience con 1-4-2-3-5-1 como la solución de prueba inicial.

13.1-2. Reconsidere el ejemplo del problema del agente viajero que se muestra en la figura 13.4.
a) Cuando se aplicó el algoritmo de subviaje inverso a este problema en la sección 13.1, la primera iteración generó un empate de cuál de los dos subviajes inversos (al invertir 3-4 o 4-5) proporcionaba la mayor reducción de la distancia del viaje; el empate se rompió de manera arbitraria a favor de la primera inversión. Determine qué hubiera pasado si se hubiese elegido la segunda de estas inversiones (al invertir 4-5).
b) Aplique el algoritmo de subviaje inverso a este problema; para ello comience con 1-2-4-5-6-7-3-1 como la solución de prueba inicial.

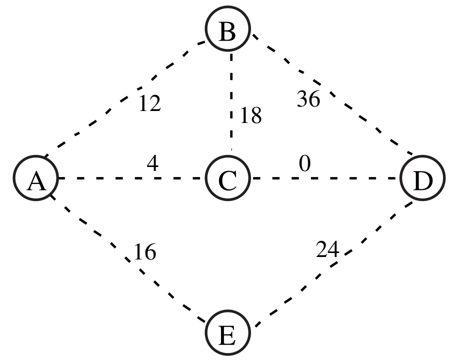
13.1-3. Considere el problema del agente viajero que se muestra a continuación, donde la ciudad 1 es la ciudad de residencia.



a) Enumere todos los viajes posibles: excluya aquellos que sólo inviertan el orden de los viajes de la lista que se presentó con anterioridad. Calcule la distancia de cada uno de estos viajes y con esto identifique el viaje óptimo.
b) Comience con 1-2-3-4-5-6-1 como la solución de prueba inicial para aplicar el algoritmo de subviaje inverso a este problema.
c) Aplique el algoritmo de subviaje inverso a este problema; para ello comience con 1-2-5-4-3-6-1 como la solución de prueba inicial.

13.2-1. Lea el artículo de referencia que describe el estudio de IO que se resume en el recuadro de aplicación que se presentó en la sección 13.2. Describa en forma breve la manera en que la búsqueda tabú se aplicó a dicho estudio. Después, elabore una lista de los beneficios financieros y de otro tipo que arrojó este estudio.

13.2.2.* Considere el problema de árbol de expansión mínima que se presenta a continuación, donde las líneas punteadas representan las ligaduras potenciales que se podrían insertar en la red, mientras que el número en seguida de cada línea punteada es el costo asociado con la inserción de esa ligadura particular.



Este problema también tiene las siguientes dos restricciones:

Restricción 1: Sólo se puede incluir una de las ligaduras AB, BC o AE.

Restricción 2: La ligadura AB se puede incluir sólo si también está incluida la ligadura BD.

Comience con la solución de prueba inicial donde las ligaduras insertadas son AB, AC, AE y CD; después, aplique a este problema el algoritmo de búsqueda tabú básico que se presentó en la sección 13.2.

13.2-3. Reconsidere el ejemplo del problema de árbol de expansión mínima restringido que se presentó en la sección 13.2 (vea la figura 13.7*a*) para obtener los datos antes de introducir las restricciones). Comience con una solución de prueba inicial diferente, a saber, la que incorpora las ligaduras AB, AD, BE y CD, y aplique de nuevo el algoritmo de búsqueda tabú básico a este problema.

13.2-4. Reconsidere el ejemplo del problema de árbol de expansión mínima no restringido que se dio en la sección 9.4. Suponga que se agregan las siguientes restricciones al problema.

Restricción 1: Es obligatorio incluir la ligadura AD o la ligadura ET.

Restricción 2: Sólo se puede incluir una de las siguientes ligaduras: AO, BC o DE.

Comience con la solución óptima del problema no restringido que se presentó al final de la sección 9.4 como la solución de prueba inicial; después, aplique el algoritmo de búsqueda tabú básico a este problema.

13.2-5. Reconsidere el problema del agente viajero que se presentó en el problema 13.1-1. Aplique de forma manual el algoritmo de búsqueda tabú básico a este problema; para ello, considere 1-5-3-2-4-1 como la solución de prueba inicial.

A **13.2-6.** Considere el problema del agente viajero con ocho ciudades cuyas ligaduras tienen las distancias asociadas que se muestran en la tabla siguiente (donde un guión indica la ausencia de una ligadura).

Ciudad	2	3	4	5	6	7	8
1	14	15	—	—	—	—	17
2		13	14	20	—	—	21
3			11	21	17	9	9
4				11	10	8	20
5					15	18	—
6						9	—
7							13

La ciudad 1 es la ciudad de residencia. Comience con cada una de las soluciones de prueba iniciales que se presentan en seguida para aplicar el algoritmo de búsqueda tabú básico del IOR Tutorial a este problema. En cada caso, cuente el número de veces que el algoritmo hace un movimiento sin mejora. También identifique los movimientos tabú que se realizan de cualquier modo porque generan la mejor solución de prueba encontrada hasta ese momento.

***a*)** Use 1-2-3-4-5-6-7-8-1 como la solución de prueba inicial.
***b*)** Use 1-2-5-6-7-4-8-3-1 como la solución de prueba inicial.
***c*)** Use 1-3-2-5-6-4-7-8-1 como la solución de prueba inicial.

A **13.2-7.** Considere el problema del agente viajero con 10 ciudades cuyas ligaduras tienen las distancias asociadas que se muestran en la tabla siguiente:

Ciudad	2	3	4	5	6	7	8	9	10
1	13	25	15	21	9	19	18	8	15
2		26	21	29	21	31	23	16	10
3			11	18	23	28	44	34	35
4				10	13	19	34	24	29
5					12	11	37	27	36
6						10	25	14	25
7							32	23	35
8								10	16
9									14

La ciudad 1 es la ciudad de residencia. Comience con cada una de las soluciones de prueba iniciales que se presentan a continuación para aplicar el algoritmo de búsqueda tabú básico del IOR Tutorial a este problema. En cada caso, cuente el número de veces que el algoritmo hace un movimiento sin mejora. También identifique los movimientos tabú que se realizan de cualquier modo porque generan la mejor solución de prueba que se encontró hasta ese momento.

***a*)** Use 1-2-3-4-5-6-7-8-9-10-1 como la solución de prueba inicial.
***b*)** Use 1-3-4-5-7-6-9-8-10-2-1 como la solución de prueba inicial.
***c*)** Use 1-9-8-10-2-4-3-6-7-5-1 como la solución de prueba inicial.

13.3-1. Durante la aplicación de un algoritmo de templado simulado a cierto problema, se ha llegado a una iteración donde el valor actual de T es $T = 2$ y el valor de la función objetivo de la solución de prueba actual es 30. Esta solución de prueba tiene cuatro vecinos inmediatos cuyos valores de la función objetivo son 29, 34, 31 y 24. Para cada uno de estos vecinos inmediatos, se desea determinar la probabilidad de que la regla de la selección del movimiento lo acepte como el candidato actual a ser la próxima solución de prueba, cuando se selecciona de manera aleatoria.

***a*)** Determine esta probabilidad de cada uno de los vecinos inmediatos cuando el objetivo es *maximizar* la función objetivo.
***b*)** Determine esta probabilidad de cada uno de los vecinos inmediatos cuando el objetivo es *minimizar* la función objetivo.

A **13.3-2.** Debido a que un algoritmo de templado simulado utiliza números aleatorios, proporciona resultados un poco diferentes cada vez que se corre. En la tabla 13.5 se muestra una aplicación del algoritmo de templado simulado básico del IOR Tutorial al ejemplo del problema del agente viajero que se presentó en la figura 13.4. Comience con la misma solución de prueba inicial (1-2-3-4-5-6-7-1) y use el IOR Tutorial para aplicar este algoritmo al mismo ejemplo cinco veces más. ¿Cuántas veces se encontró de nuevo la solución óptima (1-3-5-7-6-4-2-1 o de manera equivalente, 1-2-4-6-7-5-3-1)?

13.3-3. Reconsidere el problema del agente viajero que se presentó en el problema 13.1-1. Use 1-4-2-3-5-1 como la solución de prueba inicial; además, siga las instrucciones siguientes para aplicar el algoritmo de templado simulado básico que se presentó en la sección 13.3 a este problema.

***a*)** Realice la primera iteración a mano. Siga las instrucciones dadas al inicio de la sección de problemas para obtener los números aleatorios necesarios. Muestre su trabajo, en el cual debe incluir el uso de los números aleatorios.
A ***b*)** Use el IOR Tutorial para aplicar este algoritmo. Observe el progreso del algoritmo y registre en cada iteración cuántos candidatos a ser la próxima solución de prueba se rechazan antes de que uno sea aceptado.

A **13.3-4.** Siga las instrucciones del problema 13.3-3 del problema del agente viajero descrito en el problema 13.2-6, use 1-2-3-4-5-6-7-8-1 como la solución de prueba inicial.

A **13.3-5.** Siga las instrucciones del problema 13.3-3 para resolver el problema del agente viajero descrito en el problema 13.2-7; utilice 1-9-8-10-2-4-3-6-7-5-1 como la solución de prueba inicial.

A **13.3-6.** Debido a que un algoritmo de templado simulado utiliza números aleatorios, proporciona resultados un poco diferentes cada vez que se corre. En la tabla 13.6 se muestra una aplicación del algoritmo de templado simulado básico del IOR Tutorial al ejemplo de programación no lineal que se presentó en la sección 13.1. Comience con la misma solución de prueba inicial ($x = 15.5$) y use el IOR Tutorial para aplicar este algoritmo al mismo ejemplo cinco veces más. ¿Cuál es la mejor solución que encontró en estas cinco aplicaciones? ¿Es más cercana a la solución óptima [$x = 20$ con $f(x) = 4\ 400\ 000$] que la mejor solución que se muestra en la tabla 13.6?

13.3-7 Considere el siguiente problema de programación no convexa.

$$\text{Maximizar} \quad f(x) = x^3 - 60x^2 + 900x + 100,$$

sujeta a

$$0 \leq x \leq 31.$$

***a*)** Use la primera y segunda derivadas de $f(x)$ para determinar los puntos críticos (junto con los puntos extremos de la región factible) donde x es un máximo o un mínimo local.
***b*)** Bosqueje en forma manual la gráfica de $f(x)$ a lo largo de la región factible.
***c*)** Use $x = 15.5$ como la solución de prueba inicial y realice en forma manual la primera iteración del algoritmo de templado simulado que se presentó en la sección 13.3. Siga las instrucciones dadas al inicio de la sección de problemas para obtener los números aleatorios necesarios. Muestre su trabajo, incluyendo el uso de los números aleatorios.
A ***d*)** Use el IOR Tutorial para aplicar este algoritmo con una solución de prueba inicial de $x = 15.5$. Observe el progreso del algoritmo y registre para cada iteración cuántos candidatos a ser la próxima solución de prueba se rechazan antes de que uno sea aceptado. También cuente el número de iteraciones en las que se acepta un movimiento sin mejora.

13.3-8. Considere el ejemplo de un problema de programación no convexa que se presentó en la sección 12.10 y que se muestra en la figura 12.18.

a) Use $x = 2.5$ como la solución de prueba inicial y realice en forma manual la primera iteración del algoritmo de templado simulado que se presentó en la sección 13.3. Siga las instrucciones dadas al inicio de la sección de problemas para obtener los números aleatorios necesarios. Muestre su trabajo, que debe incluir el uso de los números aleatorios.

A ***b***) Use el IOR Tutorial para aplicar este algoritmo con una solución de prueba inicial de $x = 2.5$. Observe el progreso del algoritmo y registre en cada iteración cuántos candidatos a ser la próxima solución de prueba se rechazan antes de que uno sea aceptado. También cuente el número de iteraciones en las que se acepta un movimiento sin mejora.

A **13.3-9.** Siga las instrucciones del problema 13.3-8 para el siguiente problema de programación no convexa, con una solución de prueba inicial de $x = 25$.

$$\text{Maximizar} \quad f(x) = x^6 - 140x^5 + 7\,000x^4 - 160\,000x^3 + 1\,600\,000x^2 - 5\,000\,000x,$$

sujeto a

$$0 \leq x \leq 50.$$

A **13.3-10.** Siga las instrucciones del problema 13.3-8 para resolver el siguiente problema de programación no convexa, con una solución de prueba inicial de $(x_1, x_2) = (18, 25)$.

$$\text{Maximizar} \quad f(x_1, x_2) = x_1^5 - 81x_1^4 + 2\,330x_1^3 - 28\,750x_1^2 + 150\,000x_1 + 0.5x_2^5 - 65x_2^4 + 2\,950x_2^3 - 53\,500x_2^2 + 305\,000x_2,$$

sujeto a

$$x_1 + 2x_2 \leq 110$$
$$3x_1 + x_2 \leq 120$$

y

$$0 \leq x_1 \leq 36, \quad 0 \leq x_2 \leq 50.$$

13.4-1. Para cada una de las siguientes parejas de padres genere sus dos hijos mediante la aplicación del algoritmo genético básico que se presentó en la sección 13.4 para un problema de programación no lineal entera que involucra una sola variable x, la cual está restringida a valores enteros en el intervalo $0 \leq x \leq 63$. (Siga las instrucciones dadas al principio de la sección de problemas para obtener los números aleatorios necesarios y después muestre el uso que les dio a estos números aleatorios.)

a) Los padres son 010011 y 100101.
b) Los padres son 000010 y 001101.
c) Los padres son 100000 y 101000.

13.4.2.* Considere un problema del agente viajero con ocho ciudades (ciudades 1, 2, . . ., 8) donde la ciudad 1 es la ciudad de residencia y existen ligaduras entre cada par de ciudades. En el caso de cada una de las siguientes parejas de padres genere sus dos hijos mediante la aplicación del algoritmo genético básico que se presentó en la sección 13.4. (Siga las instrucciones dadas al principio de la sección de problemas para obtener los números aleatorios necesarios y después muestre el uso que dio a estos números aleatorios.)

a) Los padres son 1-2-3-4-7-6-5-8-1 y 1-5-3-6-7-8-2-4-1.
b) Los padres son 1-6-4-7-3-8-2-5-1 y 1-2-5-3-6-8-4-7-1.
c) Los padres son 1-5-7-4-6-2-3-8-1 y 1-3-7-2-5-6-8-4-1

A **13.4-3.** En la tabla 13.7 se muestra la aplicación del algoritmo genético básico que se describió en la sección 13.4 a un ejemplo de programación no lineal entera durante el paso inicial y la primera iteración.

a) Use el IOR Tutorial para aplicar el mismo algoritmo al mismo ejemplo, hasta la terminación del algoritmo y a partir de otra población inicial que seleccione en forma aleatoria. ¿Esta aplicación obtiene de nuevo la solución óptima ($x = 20$), como se encontró durante la primera iteración en la tabla 13.7?
b) Debido a que un algoritmo genético utiliza números aleatorios, proporciona resultados un poco diferentes cada vez que se corre. Use el IOR Tutorial para aplicar el algoritmo genético básico descrito en la sección 13.4 a este mismo ejemplo cinco veces más. ¿Cuántas veces encuentra de nuevo la solución óptima ($x = 20$)?

13.4-4. Reconsidere el problema de programación no convexa que se presentó en el problema 13.3-7. Suponga ahora que la variable x está restringida a valores enteros.

a) Ejecute a mano el paso inicial y la primera iteración del algoritmo genético básico que se presentó en la sección 13.4. Siga las instrucciones que se dieron al principio de la sección de problemas para obtener los números aleatorios necesarios. Muestre su trabajo, el cual debe incluir el uso de los números aleatorios.

A ***b***) Use el IOR Tutorial para aplicar este algoritmo. Observe el progreso del algoritmo y registre el número de veces que una pareja de padres produce un hijo cuya aptitud es mejor que la de ambos padres. También cuente el número de iteraciones donde la mejor solución que haya encontrado es mejor que cualquiera de las previas.

A **13.4-5.** Siga las instrucciones del problema 13.4-4 para resolver el problema de programación no convexa que se describió en el problema 13.3-9, cuando la variable x está restringida a valores enteros.

A **13.4-6.** Siga las instrucciones del problema 13.4-4 para resolver el problema de programación no convexa que se describió en el problema 13.3-10, cuando las dos variables x_1 y x_2 están restringidas a valores enteros.

A **13.4-7.** En la tabla 13.9 se muestra la aplicación del algoritmo genético básico que se describió en la sección 13.4 al ejemplo del problema del agente viajero que se muestra en la figura 13.4 durante el paso inicial y la primera iteración del algoritmo.

a) Use el IOR Tutorial para aplicar el mismo algoritmo al mismo ejemplo, hasta la terminación del algoritmo y a partir de otra población inicial que seleccione en forma aleatoria. ¿Esta aplicación obtiene la solución óptima (1-3-5-7-6-4-2-1 o, la equivalente, 1-2-4-6-7-5-3-1)?
b) Debido a que un algoritmo genético utiliza números aleatorios, proporciona resultados un poco diferentes cada vez que se corre. Use el IOR Tutorial para aplicar el algoritmo genético básico que se describió en la sección 13.4 a este mismo ejemplo cinco veces más. ¿Cuántas veces encuentra de nuevo la solución óptima?

13.4-8. Reconsidere el problema del agente viajero que se presentó en el problema 13.1-1.

a) Ejecute a mano el paso inicial y la primera iteración del algoritmo genético básico que se presentó en la sección 13.4. Siga las instrucciones que se dieron al principio de la sección de problemas para obtener los números aleatorios necesarios. Muestre su trabajo, incluyendo el uso de los números aleatorios.

A ***b)*** Use el IOR Tutorial para aplicar este algoritmo. Observe el progreso del algoritmo y registre el número de veces que una pareja de padres produce un hijo cuyo viaje tiene una distancia menor que la de los dos padres. También cuente el número de iteraciones donde la mejor solución que haya encontrado tiene una distancia menor que cualquiera de las previas.

A **13.4.9.** Siga las instrucciones del problema 13.4-8 para solucionar el problema del agente viajero que se describe en el problema 13.2-6.

A **13.4.10.** Siga las instrucciones del problema 13.4-8 para resolver el problema del agente viajero que se describe en el problema 13.2-7.

A **13.5-1.** Use el IOR Tutorial para aplicar el algoritmo básico de las tres metaheurísticas que se presentó en este capítulo al problema del agente viajero que se describe en el problema 13.2-6. (Use 1-2-3-4-5-6-7-8-1 como la solución de prueba inicial para los algoritmos de búsqueda tabú y templado simulado.) ¿Cuál metaheurística proporciona la mejor solución para este problema particular?

A **13.5-2.** Use el IOR Tutorial para aplicar el algoritmo básico de las tres metaheurísticas que se presentó en este capítulo al problema del agente viajero que se describió en el problema 13.2-7. (Use 1-2-3-4-5-6-7-8-9-10-1 como la solución de prueba inicial para los algoritmos de búsqueda tabú y templado simulado.) ¿Cuál metaheurística proporciona la mejor solución para este problema particular?

CAPÍTULO 14

Teoría de juegos

La vida está llena de conflictos y competencia. Los numerosos ejemplos que involucran adversarios en conflicto incluyen juegos de mesa, combates militares, campañas políticas, competencias deportivas, campañas de publicidad y comercialización en las competencias de empresas, entre otros. Una característica básica en muchas de estas situaciones es que el resultado final depende, en primer lugar, de la combinación de estrategias seleccionadas por los adversarios. La teoría de juegos es una teoría matemática que estudia las características generales de situaciones competitivas de manera formal y abstracta. Además, otorga una importancia especial a los procesos de toma de decisiones de los adversarios.

Debido a que los escenarios de competencia son algo muy común en la actualidad, la teoría de juegos tiene aplicaciones en una gran variedad de áreas, dentro de las cuales se incluyen los negocios y la economía. Por ejemplo, la referencia seleccionada 3 presenta varias aplicaciones de negocios de la teoría de juegos y la referencia seleccionada 1 se enfoca en sus aplicaciones a la economía. El premio Nobel en Economía de 1994 le fue otorgado a John F. Nash, Jr. (cuya biografía se relata en la película *A Beautiful Mind*), John C. Harsanyi y Reinhard Selton por su análisis de equilibrio en la teoría de juegos no cooperativos. Tiempo después, en 2005, Robert J. Aumann y Thomas C. Schelling ganaron el premio Nobel en la misma disciplina por la mejora que aportaron a la comprensión del conflicto y la cooperación a través del análisis de la teoría de juegos.

Como se analiza de manera breve en la sección 14.6, la investigación sobre teoría de juegos continúa con el sondeo de las situaciones competitivas de tipo complicado. No obstante, este capítulo se aboca al caso más sencillo conocido como **juegos de dos personas y suma cero**. Como su nombre lo indica, en estos juegos participan sólo dos adversarios o *jugadores* (que pueden ser ejércitos, equipos, empresas, etc.). Son llamados juegos de *suma cero* porque un jugador gana lo que el otro pierde, de manera que la suma de sus ganancias netas es cero.

Después de introducir el modelo básico de los juegos de dos personas y suma cero en la sección 14.1, en las cuatro secciones siguientes se describen e ilustran distintos enfoques para resolver este tipo de juegos. El capítulo concluye con la mención de otro tipo de situaciones competitivas que se estudian en otras ramas de la teoría de juegos.

14.1 FORMULACIÓN DE JUEGOS DE DOS PERSONAS Y SUMA CERO

Para ilustrar las características básicas de un modelo de juegos de dos personas de suma cero, considere el juego llamado *pares y nones*. Éste consiste nada más en que los dos jugadores muestran al mismo tiempo uno o dos dedos. Si el número total de dedos mostrados por ambos jugadores es par, el jugador que apuesta a pares (por ejemplo, el jugador 1) gana la apuesta (digamos 1 dólar) al jugador que elige nones (jugador 2). Si el número de dedos es impar, el jugador 1 paga 1 dólar al jugador 2. Entonces, cada jugador tiene dos *estrategias*: mostrar uno o dos dedos. El pago en dólares que resulta para el jugador 1 se muestra en una *matriz de pagos* en la tabla 14.1.

TABLA 14.1 Matriz de pagos del juego de pares y nones

		Jugador 2	
Estrategia		**1**	**2**
Jugador 1	1	1	−1
	2	−1	1

En general, un juego de dos personas se caracteriza por

1. Las estrategias del jugador 1.
2. Las estrategias del jugador 2.
3. La matriz de pagos.

Antes de iniciar el juego, cada jugador conoce las estrategias de que dispone, las que tiene su oponente y la matriz de pagos. Una jugada real consiste en que los dos jugadores elijan al mismo tiempo una estrategia sin saber cuál es la elección de su oponente.

Una estrategia puede constituir una acción sencilla, como mostrar un número par o non de dedos en el juego de pares y nones. Por otro lado, en juegos más complicados que llevan en sí una serie de movimientos, una **estrategia** es una regla predeterminada que especifica por completo cómo se intenta responder a cada circunstancia posible en cada etapa del juego. Por ejemplo, una estrategia de un jugador de ajedrez indica cómo hacer el siguiente movimiento ante *todas* las posiciones posibles en el tablero, por lo que el número total de estrategias posibles sería astronómico. Las aplicaciones de la teoría de juegos involucran situaciones competitivas mucho menos complicadas que el ajedrez, pero las estrategias que se manejan pueden llegar a ser bastante complejas.

Por lo general, la **matriz de pagos** muestra la ganancia (positiva o negativa) del jugador 1, que resultaría con cada combinación de estrategias de los dos jugadores. Se presenta sólo la matriz del jugador 1, puesto que la del jugador 2 es el negativo de ésta, debido a la naturaleza de suma cero del juego.

Los elementos de la matriz de pagos pueden expresar cualquier tipo de unidades, como dólares, siempre que representen con exactitud la *utilidad* del jugador 1 ante el resultado correspondiente. Debe hacerse hincapié en que la utilidad no necesariamente es proporcional a la cantidad de dinero (o cualquier otro bien) cuando se manejan cantidades grandes. Por ejemplo, 2 millones de dólares (después de impuestos) pueden tener un valor mucho menor que "el doble" del valor que representa 1 millón de dólares para una persona pobre. En otras palabras, si a una persona se le da a elegir entre: 1) recibir, con 50% de posibilidades, 2 millones o nada y 2) recibir 1 millón con seguridad, una persona pobre tal vez preferiría la última oferta. En otras palabras, el resultado que corresponde a un elemento con valor 2 en una matriz de pagos debe "valer el doble" para el jugador 1 que el resultado correspondiente a un elemento de 1. Así, dada la elección, debe serle indiferente 50% de posibilidades de recibir el primer resultado (en lugar de nada) y recibir en definitiva el último resultado.[1]

Un objetivo primordial de la teoría de juegos es desarrollar criterios racionales para seleccionar una estrategia, los cuales implican dos supuestos importantes:

1. *Ambos* jugadores son *racionales*.
2. *Ambos* jugadores eligen sus estrategias sólo para *promover su propio bienestar* (sin compasión para el oponente).

La teoría de juegos se contrapone al *análisis de decisión* (vea el capítulo 15), en donde se hace el supuesto de que el tomador de decisiones está jugando un juego contra un oponente pasivo —la naturaleza— que elige sus estrategias de alguna manera aleatoria.

[1] Vea en la sección 15.6 una presentación más completa del concepto de utilidad.

En este capítulo se desarrollará el criterio estándar de la teoría de juegos para elegir las estrategias mediante ejemplos ilustrativos. En particular, al final de la siguiente sección se describe la forma en que la teoría de juegos dice cómo debe jugarse el juego de pares y nones. (Los problemas 14.3-1, 14.4-1 y 14.5-1 lo invitan también a aplicar las técnicas desarrolladas en este capítulo para encontrar la forma óptima de acometer este juego.) Además, la sección siguiente presenta un ejemplo prototipo que ilustra la formulación de un juego y su solución en algunas situaciones sencillas. Después, en la sección 14.3, se desarrollará una variación más complicada de este juego para obtener un criterio más general. En las secciones 14.4 y 14.5 se describe un procedimiento gráfico y una formulación de programación lineal para juegos de este tipo.

14.2 SOLUCIÓN DE JUEGOS SENCILLOS: EJEMPLO PROTOTIPO

Dos políticos contienden entre sí por un lugar en el Senado de Estados Unidos. En este momento elaboran sus planes de campaña para los dos últimos días antes de las elecciones; se espera que dichos días sean cruciales debido a que están muy próximos al día de la votación. Por esta circunstancia, ambos quieren emplearlos para hacer campaña en dos ciudades importantes: Bigtown y Megalópolis. Para evitar pérdidas de tiempo, planean viajar en la noche y pasar un día completo en cada ciudad o dos días en sólo una de ellas. Como deben hacer los arreglos necesarios por adelantado, ninguno de los dos conocerá lo que su oponente tiene planeado hacer hasta después de concretar sus propios planes. Cada político tiene un jefe de campaña en cada ciudad para asesorarlo sobre el efecto que tendrán (en términos de votos ganados o perdidos) las combinaciones posibles de los días dedicados a cada ciudad por ellos o por sus oponentes. Por tanto, quieren emplear esta información para elegir su mejor estrategia para estos dos días.

Formulación como un juego de dos personas y suma cero

Para formular este problema como un juego de dos personas y suma cero se deben identificar los *dos jugadores* (obviamente, los dos políticos), las *estrategias* de cada uno de ellos y la *matriz de pagos*.

Según la forma en que se estableció el problema, cada jugador tiene tres estrategias:

Estrategia 1 = pasar un día en cada ciudad.
Estrategia 2 = pasar los dos días en Bigtown.
Estrategia 3 = pasar los dos días en Megalópolis.

Por el contrario, las estrategias serían más complicadas en una situación diferente en la que cada político supiera en dónde pasará su oponente el primer día antes de concluir sus propios planes para el segundo día. En ese caso, una estrategia normal sería: pasar el primer día en Bigtown; si el oponente también pasa el día en Bigtown, entonces quedarse el segundo día en esa ciudad; sin embargo, si el oponente pasa el primer día en Megalópolis, entonces sería necesario pasar el segundo día en ese lugar. Habría ocho estrategias de este tipo, una para cada combinación de las dos posibilidades para el primer día, las dos para el primer día del oponente y las dos alternativas para el segundo día.

Cada elemento de la matriz de pagos del jugador 1 representa la *utilidad* para ese jugador (o la utilidad negativa para el jugador 2) de los resultados que se obtienen cuando los dos jugadores emplean las estrategias correspondientes. Desde el punto de vista de los políticos, el objetivo es *ganar votos* y cada voto adicional (antes de conocer el resultado de las elecciones) tiene el mismo valor para él. Entonces, los elementos apropiados de la matriz de pagos se darán en términos del *total neto de votos ganados* a su oponente (esto es, la suma de la cantidad neta de cambios de votos en las dos ciudades) que resulte de estos dos días de campaña. Con unidades de 1 000 votos, esta formulación se resume en la tabla 14.2. La teoría de juegos supone que ambos jugadores usan la misma formulación (incluso los mismos pagos para el jugador 1) para elegir sus estrategias.

Sin embargo, también debe hacerse notar que esta matriz de pagos *no* sería apropiada si se contara con información adicional. En particular, suponga que se conoce con exactitud cuáles son los planes de votación de la población dos días antes de las elecciones, de manera que cada político sabe la cantidad neta de votos (positiva o negativa) que necesita cambiar a su favor durante los dos últimos días de campaña para ganar. Entonces, lo único significativo de los datos que se pro-

TABLA 14.2 Formulación de la matriz de pagos del problema de la campaña política

Estrategia		Cantidad neta en unidades de votos ganados por el político 1 (cada unidad equivale a 1 000 votos) Político 2		
		1	2	3
	1			
Político 1	2			
	3			

porcionan en la tabla 14.2 sería que indican cuál es el político que ganaría las elecciones con cada combinación de estrategias. Como la meta final es ganar la elección y como el tamaño de la mayoría no tiene consecuencias, los elementos de utilidad de la matriz deben ser constantes positivas (digamos $+1$) cuando el político 1 gana y -1 cuando pierde. Aun cuando sólo se pudiera determinar una *probabilidad* de ganar para cada combinación de estrategias, los elementos apropiados de la matriz serían la probabilidad de ganar menos la probabilidad de perder, pues de esta forma representarían las utilidades *esperadas*. Sin embargo, casi nunca se dispone de datos con suficiente exactitud como para hacer ese tipo de determinaciones, por lo que este ejemplo utiliza los miles de votos netos totales ganados por el político 1 como elementos de la matriz de pagos.

Cuando se utiliza la forma dada en la tabla 14.2 se obtienen tres conjuntos de datos alternativos para la matriz de pagos, a fin de ilustrar cómo se resuelven tres tipos distintos de juegos.

Variación 1 del ejemplo

Si la tabla 14.3 representa la matriz de pagos del jugador l (político 1), ¿qué estrategia debe elegir cada uno?

Esta situación es bastante especial; en ella se puede obtener la respuesta con sólo aplicar el concepto de **estrategia dominada** para eliminar una serie de estrategias inferiores hasta que quede sólo una que se pueda elegir.

> Una estrategia es **dominada** por una segunda estrategia si esta última es *siempre al menos tan buena* (y algunas veces mejor) como la primera, sin que importe lo que haga el oponente. Una estrategia dominada se puede eliminar de inmediato para consideraciones posteriores.

En principio, la tabla 14.3 no incluye las estrategias dominadas del jugador 2; pero para el jugador 1, la estrategia 3 está dominada por la estrategia 1, ya que tiene pagos más altos ($1 > 0$, $2 > 1$, $4 > -1$), independientemente de lo que haga el jugador 2. Al eliminar la estrategia 3 se obtiene la siguiente matriz de pagos reducida:

	1	2	3
1	1	2	4
2	1	0	5

TABLA 14.3 Matriz de pagos de la variación 1 del problema de la campaña política

Estrategia		Jugador 2		
		1	2	3
	1	1	2	4
Jugador 1	2	1	0	5
	3	0	1	−1

Como se supone que ambos jugadores son racionales, también el jugador 2 puede deducir que el jugador 1 sólo dispone de estas dos estrategias. Entonces, ahora el jugador 2 *tiene* una estrategia dominada: la estrategia 3, que está dominada tanto por la estrategia 1 como por la 2 puesto que siempre tienen menores pérdidas (pagos al jugador 1) en esta matriz de pagos reducida (para la estrategia 1: $1 < 4$, $1 < 5$; para la estrategia 2: $2 < 4$, $0 < 5$). Cuando se elimina esta estrategia se obtiene

	1	2
1	1	2
2	1	0

En este punto, la estrategia 2 del jugador 1 se convierte en dominada por la estrategia 1, puesto que esta última es mejor en la columna 2 ($2 > 0$) y es igual en la columna 1 ($1 = 1$). Si se elimina esta estrategia dominada se llega a

	1	2
1	1	2

Ahora la estrategia 2 del jugador 2 está dominada por la estrategia 1 ($1 < 2$), por lo que debe eliminarse la estrategia 2.

En consecuencia, ambos jugadores deberán elegir su estrategia 1. Con esta solución, el jugador 1 recibirá un pago de 1 por parte del jugador 2 (esto es, el político 1 ganará 1 000 votos al político 2).

Si usted quiere ver **otro ejemplo** donde se muestre cómo resolver un juego utilizando el concepto de estrategias dominantes, en la sección Worked Examples del sitio en internet de este libro se proporciona uno.

En general, el pago para el jugador 1 cuando ambos jugadores operan de manera óptima recibe el nombre de **valor del juego**. Se dice que se trata de un **juego justo** cuando el juego tiene valor 0. Como este juego en particular tiene valor 1, *no* es un juego justo.

El concepto de estrategia dominada es muy útil para reducir el tamaño de la matriz de pagos en cuestión y en algunos casos raros como éste puede, de hecho, identificar la solución óptima del juego. Sin embargo, casi todos los juegos requieren otro enfoque para terminar de resolverlos, como se ilustra en las dos variaciones siguientes del ejemplo.

Variación 2 del ejemplo

Ahora suponga que los datos actuales se dan en la tabla 14.4 como la matriz de pagos del jugador 1 (político 1). Este juego no tiene estrategias dominadas por lo que no es obvio qué deben hacer los jugadores. ¿Qué línea de razonamiento dice la teoría de juegos que debe usarse en este caso?

Considere al jugador 1. Si elige la estrategia 1 puede ganar 6 o perder 3. Como el jugador 2 es racional y tratará de aplicar una estrategia que lo proteja de pagos grandes al jugador 1, parece probable que si juega la estrategia 1, el jugador 1 perderá. De manera similar, si selecciona la estrategia 3 el jugador 1 puede ganar 5, pero quizá sea más probable que su oponente racional evite esta pérdida y en su lugar logre que él pierda, lo que puede ascender a 4. Por otro lado, si el jugador 1 elige la estrategia 2, tiene garantizado que no perderá y quizá gane algo. Entonces, por proporcio-

TABLA 14.4 Matriz de pagos del jugador 1 en la variación 2 del problema de la campaña política

		Jugador 2			
	Estrategia	**1**	**2**	**3**	**Mínimo**
	1	−3	−2	6	−3
Jugador 1	2	2	0	2	0 ← Valor maximin
	3	5	−2	−4	−4
	Máximo:	5	0 ↑ Valor minimax	6	

nar la *mejor garantía* (un pago de 0), la estrategia 2 parece ser la elección "racional" del jugador 1 contra su oponente racional. (Esta línea de razonamiento supone que ambos jugadores tienen aversión a arriesgar pérdidas más grandes que las necesarias, al contrario de quienes disfrutan la apuesta por un gran pago con pocas posibilidades.)

Ahora considere al jugador 2. Puede perder tanto como 5 o 6 si sigue las estrategias 1 o 3, pero está garantizado que al menos empata con la estrategia 2. Entonces, si usa el mismo razonamiento para buscar su mejor garantía contra su oponente racional, parece que la mejor elección es la estrategia 2.

Si los dos jugadores eligen la estrategia 2, el resultado es que empatan. Así, en este caso, ningún jugador mejora con su mejor garantía, pero ambos fuerzan al oponente hacia la misma posición. Aunque cada jugador deduzca la estrategia del otro, no puede explotar esta información para mejorar su posición. Se trata de un empate.

El producto final de esta línea de razonamiento es que cada jugador debe jugar de tal manera que *minimice su pérdida máxima*, siempre que el resultado de su elección no pueda ser aprovechado por su oponente para mejorar su posición. Esto se conoce como **criterio minimax**, criterio estándar que propone la teoría de juegos para elegir una estrategia. En realidad, este criterio sostiene que se debe seleccionar la mejor estrategia, aun cuando la elección fuera anunciada al oponente antes de que éste eligiera su estrategia. En términos de la matriz de pagos, implica que el *jugador 1* debe elegir aquella estrategia cuyo *pago mínimo* sea el *mayor*, mientras que el *jugador 2* debe elegir aquella cuyo *pago máximo al jugador 1* sea el *menor*. Este criterio se muestra en la tabla 14.4, donde se identifica la estrategia 2 como la *estrategia maximin* del jugador 1, y la estrategia 2 es la *estrategia minimax* del jugador 2. El pago de 0 que resulta es el valor del juego, por lo que éste es un juego justo.

Observe que en esta matriz de pagos el mismo elemento proporciona tanto el valor mínimo como el máximo. Este interesante hecho se debe a que tal elemento, por un lado, es el mínimo del renglón y, por el otro, el máximo de la columna. Esta posición de un elemento se llama **punto silla**.

El hecho de que este juego posea un punto silla es en realidad esencial para determinar cómo debe jugarse. A causa de este punto silla, ningún jugador puede aprovechar la estrategia de su oponente para mejorar su propia posición. En particular, si el jugador 2 predice o sabe que el jugador 1 empleará la estrategia 2, y cambia su plan original de usar la estrategia 2 sólo aumentará sus pérdidas. De igual manera, el jugador 1 sólo empeoraría su posición si cambiara su plan. Así, ningún jugador tiene motivos para considerar un cambio de estrategias, para quedar con ventaja respecto de su oponente, o para evitar que su oponente tenga ventaja. Entonces, ésta es una **solución estable** (llamada también *solución de equilibrio*), y cada jugador debe, exclusivamente, emplear sus respectivas estrategias maximin y minimax.

La siguiente variación ilustrará que algunos juegos no tienen punto silla y que en ese caso se requiere de un análisis más complicado.

Variación 3 del ejemplo

La información reciente sobre la campaña da como resultado la matriz de pagos final de los dos políticos (jugadores) que se muestra en la tabla 14.5. ¿Cómo debe jugarse este juego?

TABLA 14.5 Matriz de pagos del jugador 1 de la variación 3 del problema de la campaña política

		Jugador 2			
Estrategia		**1**	**2**	**3**	**Mínimo**
	1	0	−2	2	−2 ← Valor maximin
Jugador 1	2	5	4	−3	−3
	3	2	3	−4	−4
	Máximo:	5	4	2 ↑ Valor minimax	

Suponga que ambos jugadores quieren aplicar el criterio minimax igual que en la variación 2. El jugador 1 puede garantizar que no perderá más de 2 si juega la estrategia 1. De la misma manera, el jugador 2 puede asegurar que no perderá más de 2 si elige la estrategia 3.

Sin embargo, observe que, en este caso, el valor máximo (-2) y el valor mínimo (2) no coinciden. El resultado es que *no* hay *punto silla*.

¿Cuáles serán las consecuencias si ambos jugadores planean usar las estrategias mencionadas? Se puede observar que el jugador 1 ganaría 2 al jugador 2, lo que molestaría a este último. Como éste es racional, puede prever el resultado, con lo que concluiría que puede actuar mejor si gana 2 en lugar de perder 2 al jugar la estrategia 2. Como el jugador 1 también es racional, prevendría este cambio y concluiría que él también puede mejorar mucho, de –2 a 4, si cambia a la estrategia 2. Al darse cuenta de esto, el jugador 2 tomaría en cuenta regresar a la estrategia 3 para convertir la pérdida de 4 en una ganancia de 3. Este cambio posible causaría que el jugador 1 usara de nuevo la estrategia 1, después de lo cual volvería a comenzar todo el ciclo. Por lo tanto, aunque este juego se juega sólo una vez, *cualquier* elección tentativa de una estrategia deja al jugador en posición de considerar un cambio de estrategia, ya sea para tener ventaja sobre su oponente o para evitar que el oponente tenga ventaja sobre él.

En pocas palabras, la solución que se sugirió al principio —estrategia 1 para el jugador 1 y estrategia 3 para el jugador 2— es una **solución inestable**; con esto se ve la necesidad de desarrollar una solución más satisfactoria. Pero, ¿qué clase de solución debe ser ésta?

El hecho fundamental parece ser que siempre que se puede predecir la estrategia de un jugador, el oponente puede aprovechar esta información para mejorar su posición. Por lo tanto, una característica esencial de un plan racional para jugar un juego como éste es que ningún competidor pueda predecir qué estrategia usará el otro. Así, en este caso, en lugar de aplicar algún criterio conocido para determinar una sola estrategia que se usará en forma definitiva, es necesario elegir entre las estrategias aceptables de alguna manera aleatoria. Al hacer esto, ningún jugador conoce de antemano cuál de sus propias estrategias se usará, mucho menos la de su oponente.

Se plantea la misma situación con el juego de pares y nones que se presentó en la sección 14.1. La tabla de pagos para este juego que se muestra en la tabla 14.1 no tiene un punto de silla, por lo que el juego no tiene una solución estable en relación a qué estrategia (mostrar uno o dos dedos) debe seleccionar cada jugador en cada jugada. En realidad, sería una tontería que un jugador siempre mostrara el mismo número de dedos ya que entonces el oponente podría empezar a mostrar siempre el número de dedos que ganaría cada vez. Aun si la estrategia de un jugador fuera predecible de alguna forma debido a tendencias pasadas o patrones, el oponente puede aprovechar esta información para mejorar sus posibilidades de ganar. De acuerdo con la teoría de juegos, la forma racional de jugar al juego de pares y nones consiste en hacer que la elección de la estrategia sea completamente aleatoria cada vez. Esta situación se puede lograr, por ejemplo, si se lanza al aire una moneda (sin mostrar el resultado al oponente) y después se muestra, digamos, un dedo si la moneda cayó del lado del águila y dos dedos si cayó del otro lado.

Esto sugiere, en términos muy generales, el tipo de enfoque que se requiere para juegos sin punto silla. La siguiente sección presenta este enfoque en forma más completa. Dadas estas bases, en las siguientes dos secciones se desarrollarán procedimientos para encontrar la manera óptima de jugar estos juegos. Se continuará con el uso esta variación específica del problema de la campaña política con el fin de ejemplificar las ideas a medida que éstas se desarrollen.

14.3 JUEGOS CON ESTRATEGIAS MIXTAS

Cuando un juego no tiene punto silla, la teoría de juegos aconseja a cada jugador asignar una distribución de probabilidad sobre su conjunto de estrategias. Para expresar este consejo de manera matemática, sea

x_i = probabilidad de que el jugador 1 use la estrategia i ($i = 1, 2, \ldots, m$),
y_j = probabilidad de que el jugador 2 use la estrategia j ($j = 1, 2, \ldots, n$),

donde m y n son los números respectivos de estrategias disponibles. Entonces, el jugador 1 especificará su plan de juego al asignar valores a $x_1, x_2, \ldots, x_m$. Como estos valores son probabilidades, tendrán que ser no negativos y sumar 1. De igual manera, el plan del jugador 2 se describe mediante

los valores que asigne a sus variables de decisión $y_1, y_2, \ldots, y_n$. Por lo general se hace referencia a estos planes $(x_1, x_2, \ldots, x_m)$ y $(y_1, y_2, \ldots, y_n)$ con el nombre de **estrategias mixtas**, y las estrategias originales se llaman **estrategias puras**.

En el momento de jugar, es necesario que cada participante use una de sus estrategias puras, pero debe elegirla mediante algún dispositivo aleatorio para obtener una observación aleatoria que siga la distribución de probabilidad que especifica la estrategia mixta; esta observación indicará la estrategia pura que se debe usar.

A manera de ilustración, suponga que los jugadores 1 y 2 en la variación 3 del problema de la campaña política (vea la tabla 14.5) eligen las estrategias mixtas $(x_1, x_2, x_3) = (\frac{1}{2}, \frac{1}{2}, 0)$ y $(y_1, y_2, y_3) = (0, \frac{1}{2}, \frac{1}{2})$, respectivamente. Esta selección indica que el jugador 1 da igual oportunidad (probabilidad de $\frac{1}{2}$) a la elección de las estrategias (puras) 1 o 2, pero que descarta por completo la estrategia 3. De manera análoga, el jugador 2 elige al azar entre sus dos últimas estrategias puras. Para llevar a cabo la jugada, cada jugador podría tirar una moneda al aire para determinar cuál de sus dos estrategias aceptables usará.

Aunque no se cuenta con una medida de desempeño satisfactoria por completo para evaluar las estrategias mixtas, el *pago esperado* es una herramienta útil. Al aplicar la definición de valor esperado de la teoría de probabilidad, esta cantidad es

$$\text{Pago esperado para el jugador } 1 = \sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} x_i y_j,$$

donde p_{ij} es el pago si el jugador 1 usa la estrategia pura i y el jugador 2 usa la estrategia pura j. En el ejemplo de estrategias mixtas que se acaba de dar existen cuatro pagos posibles (–2, 2, 4, –3), donde cada uno ocurre con una probabilidad de $\frac{1}{4}$, el pago esperado es $\frac{1}{4}(-2 + 2 + 4 - 3) = \frac{1}{4}$. Así, esta medida de desempeño no revela nada sobre los riesgos inherentes al juego, pero indica a qué cantidad tendería el pago promedio si el juego se efectuara muchas veces.

Con esta medida, la teoría de juegos puede extender el concepto del **criterio minimax** a juegos que no tienen punto silla y que, por tanto, necesitan estrategias mixtas. En este contexto, el criterio minimax sostiene que un jugador debe elegir la estrategia mixta que *minimice la máxima pérdida que espera* para sí mismo. De manera equivalente, si se analizan los pagos (jugador 1) en lugar de las pérdidas (jugador 2), este criterio es *maximin*, es decir, *maximizar el pago esperado mínimo* para el jugador. Por *pago esperado mínimo* se entiende el pago esperado más pequeño posible que puede resultar de cualquier estrategia mixta con la que el oponente puede contar. De esta forma, la estrategia mixta del jugador 1 que es *óptima*, según este criterio, es la que proporciona la *garantía* (el mínimo pago esperado) de que es la *mejor* (máxima). (El valor de esta mejor garantía es el *valor maximin* y se denota por $\underline{v}$.) De manera similar, la estrategia *óptima* del jugador 2 es la que proporciona la *mejor garantía*, donde *mejor* significa *mínima* y *garantía* se refiere a la *máxima pérdida que espera* poder lograr con cualquiera de las estrategias mixtas del oponente. (La mejor garantía es el *valor minimax* que se denota por $\overline{v}$.)

Recuerde que cuando sólo se usan estrategias puras, resulta que los juegos que no tienen punto silla son *inestables* (sin soluciones estables). La razón esencial es que $\underline{v} < \overline{v}$, por lo que los jugadores quieren cambiar sus estrategias para mejorar su posición. De manera parecida, en los juegos con estrategias mixtas es necesario que $\underline{v} = \overline{v}$ para que la solución óptima sea *estable*. Por fortuna, según el teorema minimax de teoría de juegos, esta condición siempre se cumple para estos juegos.

> **Teorema minimax:** Si se permiten estrategias mixtas, el par de estrategias que es óptimo de acuerdo con el criterio minimax proporciona una *solución estable* con $\underline{v} = \overline{v} = v$ (el valor del juego), de manera que ninguno de los dos jugadores puede mejorar si cambia de manera unilateral su estrategia.

En la sección 14.5 se incluye una demostración de este teorema.

Aunque el concepto de estrategias mixtas se convierte en un proceso bastante intuitivo si es jugado *repetidas veces*, si requiere de una interpretación cuando sólo va a ser jugado *una vez*. En este caso, el uso de una estrategia mixta todavía implica elegir y usar *una* estrategia pura, elegida en forma aleatoria a partir de una distribución de probabilidad específica, y podría parecer más sensato pasar por alto este proceso de aleatorización y sólo escoger la "mejor" estrategia pura. Sin embargo, ya se ilustró en la variación 3 de la sección anterior que un jugador *no* puede permitir

que su oponente deduzca cuál será su estrategia (es decir, el procedimiento de solución bajo las reglas de teoría de juegos no debe identificar de manera *definitiva* la estrategia pura que se usará, cuando el juego es inestable). Todavía más, aun cuando el oponente pueda usar sólo su conocimiento de las tendencias del primer jugador para deducir probabilidades —de la estrategia pura elegida— que sean distintas a las de estrategia mixta óptima, entonces el oponente todavía puede aprovechar este conocimiento para reducir el pago esperado para el primer jugador. Así pues, la única manera de garantizar que se logre el pago esperado óptimo v, es elegir de manera aleatoria la estrategia pura que debe usarse, a partir de la distribución de probabilidad de la estrategia mixta óptima. (En la sección 20.4 se estudian los procedimientos estadísticos válidos para llevar a cabo esa selección aleatoria.)

A continuación se mostrará cómo cada jugador encuentra su estrategia mixta óptima. Se dispone de varios métodos. Uno es un procedimiento gráfico que se puede usar siempre que uno de los jugadores tenga sólo dos estrategias puras (no dominadas); este enfoque se describe en la siguiente sección. Cuando se trata de juegos más grandes, el método que más se emplea consiste en transformar el problema en uno de programación lineal que se puede resolver por el método símplex en una computadora; en la sección 14.5 se analiza este enfoque.

14.4 PROCEDIMIENTO DE SOLUCIÓN GRÁFICO

Considere cualquier juego con estrategias mixtas tal que después de eliminar las estrategias dominadas, uno de los jugadores tiene sólo dos estrategias puras. Para ser específico, sea éste el jugador 1. Como sus estrategias mixtas son (x_1, x_2) y $x_2 = 1 - x_1$, nada más debe obtener el valor óptimo de x_1. Sin embargo, resulta directo y sencillo hacer la gráfica del pago esperado como una función de x_1, para cada una de las estrategias de su oponente. Esta gráfica se puede usar para identificar el punto que maximiza el mínimo pago esperado. También es posible identificar en ella la estrategia mixta minimax del oponente.

Para ilustrar este procedimiento, considere la variación 3 del problema de la campaña política (vea la tabla 14.5). Observe que la tercera estrategia pura del jugador 1 está dominada por la segunda, por lo que la matriz de pagos se puede reducir a la forma dada en la tabla 14.6. Entonces, para cada estrategia pura de que dispone el jugador 2, el pago esperado para el jugador 1 será:

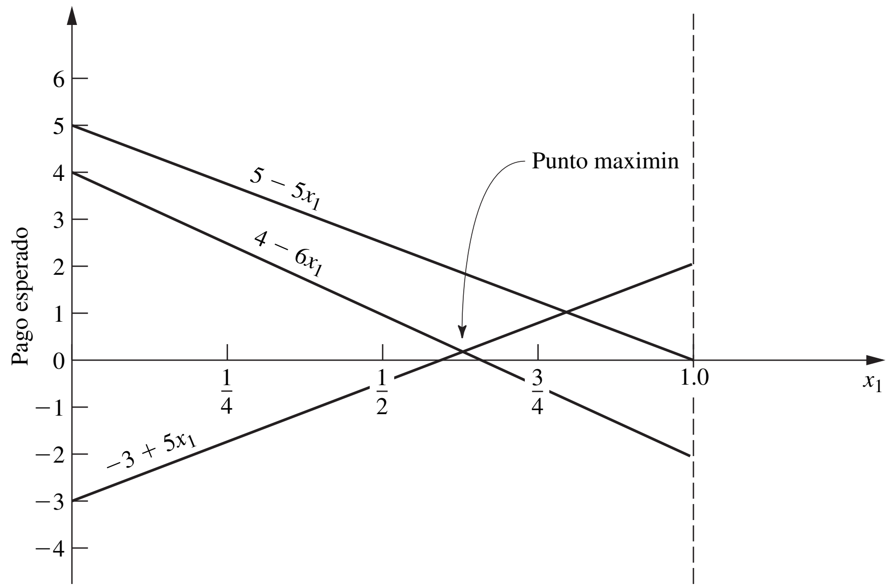
(y_1, y_2, y_3)	Pago esperado
(1, 0, 0)	$0x_1 + 5(1 - x_1) = 5 - 5x_1$
(0, 1, 0)	$-2x_1 + 4(1 - x_1) = 4 - 6x_1$
(0, 0, 1)	$2x_1 - 3(1 - x_1) = -3 + 5x_1$

A continuación se grafican estas rectas del pago esperado, como se muestra en la figura 14.1. Para cualquier valor dado de x_1 y de (y_1, y_2, y_3), el pago esperado será el promedio ponderado apropiado de los puntos correspondientes a estas tres rectas. En particular,

$$\text{Pago esperado para el jugador } 1 = y_1(5 - 5x_1) + y_2(4 - 6x_1) + y_3(-3 + 5x_1).$$

TABLA 14.6 Tabla de pagos reducida del jugador 1 de la variación 3 del problema de la campaña política

			Jugador 2		
		Probabilidad	y_1	y_2	y_3
	Probabilidad	Estrategia pura	1	2	3
Jugador 1	x_1	1	0	−2	2
	$1 - x_1$	2	5	4	−3



■ **FIGURA 14.1**
Procedimiento gráfico para la solución de los juegos.

Recuerde que el jugador 2 quiere minimizar este pago esperado para el jugador 1. Dado x_1, el jugador 2 puede minimizar este pago esperado si elige la estrategia pura que corresponde a la recta "inferior" de esa x_1 en la figura 14.1 (ya sea $-3 + 5x_1$ o $4 - 6x_1$, pero no $5 - 5x_1$). Según el criterio minimax (o maximin), el jugador 1 quiere maximizar este pago esperado mínimo. En consecuencia, debe elegir el valor de x_1 para el que la recta inferior tenga su mayor valor, es decir, el valor en el que las rectas $(-3 + 5x_1)$ y $(4 - 6x_1)$ se cruzan, de donde se obtiene un pago esperado de

$$\underline{v} = v = \max_{0 \le x_1 \le 1} \{\min\{-3 + 5x_1, 4 - 6x_1\}\}.$$

Para encontrar en forma algebraica el valor óptimo de x_1 en la intersección de las dos rectas $-3 + 5x_1$ y $4 - 6x_1$, se establece

$$-3 + 5x_1 = 4 - 6x_1,$$

de donde se obtiene $x_1 = \frac{7}{11}$. Así, $(x_1, x_2) = (\frac{7}{11}, \frac{4}{11})$ es la *estrategia mixta óptima* del jugador 1, y

$$\underline{v} = v = -3 + 5\left(\frac{7}{11}\right) = \frac{2}{11}$$

es el valor del juego.

Para encontrar la estrategia mixta óptima correspondiente del jugador 2, el razonamiento es el siguiente. De acuerdo con la definición del valor minimax $\overline{v}$ y el teorema minimax, el pago esperado que se obtiene con esta estrategia $(y_1, y_2, y_3) = (y_1^*, y_2^*, y_3^*)$ tendrá que satisfacer la condición

$$y_1^*(5 - 5x_1) + y_2^*(4 - 6x_1) + y_3^*(-3 + 5x_1) \le \overline{v} = v = \frac{2}{11}$$

para todos los valores de x_1 ($0 \le x_1 \le 1$). Más aún, cuando el jugador 1 juega de manera óptima —esto es, $x_1 = \frac{7}{11}$—, esta desigualdad será una igualdad —por el teorema minimax—, de manera que

$$\frac{20}{11}y_1^* + \frac{2}{11}y_2^* + \frac{2}{11}y_3^* = v = \frac{2}{11}.$$

Como (y_1, y_2, y_3) es una distribución de probabilidad, también se sabe que

$$y_1^* + y_2^* + y_3^* = 1.$$

Por lo tanto, $y_1^* = 0$, puesto que $y_1^* > 0$ violaría la penúltima ecuación; es decir, en la gráfica, el pago esperado en el punto $x_1 = \frac{7}{11}$ estaría por encima del punto maximin. (En general, a cualquier recta que no pasa por el punto maximin se le debe asignar un peso de cero para evitar que el pago esperado tenga un valor más alto que este punto.)

Entonces,

$$y_2^* (4 - 6x_1) + y_3^* (-3 + 5x_1) \begin{cases} \leq \dfrac{2}{11} & \text{para } \ 0 \leq x_1 \leq 1, \\ = \dfrac{2}{11} & \text{para } \ x_1 = \dfrac{7}{11}. \end{cases}$$

Pero y_2^* y y_3^* son números y entonces el lado derecho es la ecuación de una recta, lo cual es un peso ponderado fijo de las dos rectas "inferiores" de la gráfica. Como la ordenada de esta recta debe ser igual a $\frac{2}{11}$ en el punto $x_1 = \frac{7}{11}$ y como nunca debe exceder $\frac{2}{11}$, la recta es horizontal por necesidad. (Esta conclusión siempre es válida a menos que el valor óptimo de x_1 sea 0 o 1, en cuyo caso el jugador 2 también debe usar una sola estrategia pura.) Por tanto,

$$y_2^*(4 - 6x_1) + y_3^*(-3 + 5x_1) = \frac{2}{11}, \qquad \text{para } 0 \leq x_1 \leq 1.$$

Entonces, para obtener y_2^* y y_3^* se seleccionan dos valores de x_1 (como 0 y 1) y se resuelven las dos ecuaciones simultáneas obtenidas. Así,

$$4y_2^* - 3y_3^* = \frac{2}{11},$$

$$-2y_2^* + 2y_3^* = \frac{2}{11},$$

de donde $y_2^* = \frac{5}{11}$ y $y_3^* = \frac{6}{11}$. Por tanto, la *estrategia mixta óptima* del jugador 2 es $(y_1, y_2, y_3) = (0, \frac{5}{11}, \frac{6}{11})$.

Si en algún otro problema ocurre que más de dos rectas pasan por el punto maximin, de manera que más de dos valores y_j^* pueden ser mayores que cero, entonces habría muchos empates en el caso de la estrategia mixta óptima del jugador 2. Una de estas estrategias se puede identificar si se igualan en forma arbitraria todas menos dos de estos valores y_j^* a cero y se obtienen las dos restantes como se acaba de describir. En el caso de esas dos y_j^*, las rectas asociadas deben tener una pendiente positiva en un caso y una pendiente negativa en el otro.

Aunque este procedimiento gráfico se ejemplificó para un problema en particular, en esencia el mismo razonamiento se puede usar para resolver cualquier juego con estrategias mixtas que tenga sólo dos estrategias puras no dominadas de uno de los jugadores. En la sección Worked Examples del sitio en internet de este libro se proporciona **otro ejemplo** donde, en este caso, es el jugador 2 quien tiene sólo dos estrategias no dominadas, de forma que el procedimiento de solución gráfica se aplica en un inicio desde el punto de vista de este jugador.

14.5 SOLUCIÓN MEDIANTE PROGRAMACIÓN LINEAL

Cualquier juego de estrategias mixtas se puede resolver en forma muy sencilla si se lo transforma en un problema de programación lineal. Como se verá, esta transformación requiere apenas un poco más que la aplicación del teorema minimax y el uso de la definición de valor maximin $\underline{v}$ y valor minimax $\overline{v}$.

Primero, considere cómo se encuentra la estrategia mixta del jugador 1. Como se indicó en la sección 14.3,

$$\text{Pago esperado para el jugador 1} = \sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} x_i y_j$$

y la estrategia $(x_1, x_2, \ldots, x_m)$ es óptima si

$$\sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} x_i y_j \geq \underline{v} = v$$

para todas las estrategias del oponente ($y_1, y_2, \ldots, y_n$). Entonces, esta desigualdad se debe cumplir, por ejemplo, para cada una de las estrategias puras del jugador 2, es decir, para cada una de las estrategias ($y_1, y_2, \ldots, y_n$) donde $y_j = 1$ y el resto es igual a 0. Al sustituir estos valores en la desigualdad se obtiene

$$\sum_{i=1}^{m} p_{ij}x_i \geq v \qquad \text{para } j = 1, 2, \ldots, n,$$

de manera que la desigualdad *implica* este conjunto de n desigualdades. Aún más, este conjunto de n desigualdades *implica* la desigualdad original (escrita de otra forma)

$$\sum_{j=1}^{n} y_j\left(\sum_{i=1}^{m} p_{ij}x_i\right) \geq \sum_{j=1}^{n} y_j v = v,$$

porque

$$\sum_{j=1}^{n} y_j = 1.$$

Como la implicación va en ambas direcciones, se concluye que imponer este conjunto de n desigualdades lineales es *equivalente* a requerir que la desigualdad original se cumpla para todas las estrategias ($y_1, y_2, \ldots, y_n$). Pero estas n desigualdades son restricciones válidas en programación lineal, como lo son las restricciones adicionales

$$x_1 + x_2 + \ldots + x_m = 1$$
$$x_i \geq 0, \quad \text{para } i = 1, 2, \ldots, m$$

que se necesitan para asegurar que las x_i sean probabilidades. Por esta razón, cualquier solución ($x_1, x_2, \ldots, x_m$) que satisfaga este conjunto completo de restricciones de programación lineal es la estrategia mixta óptima deseada.

En consecuencia, el problema de encontrar una estrategia mixta óptima se ha reducido a encontrar una solución factible para un problema de programación lineal, lo que se puede hacer según se describió en el capítulo 4. Las dos dificultades que quedan por resolver son que: 1) se desconoce v, y 2) el problema de programación lineal no tiene función objetivo. Por fortuna, ambos obstáculos se pueden salvar al mismo tiempo si se sustituye la constante desconocida v por la variable x_{m+1} y después se maximiza x_{m+1}, de manera que en forma automática x_{m+1} será igual a v (por definición) en la solución óptima del problema de programación lineal.

La formulación de la programación lineal

Para resumir, el jugador 1 encontrará su estrategia mixta óptima al emplear el método símplex para resolver el problema de programación lineal:

$$\text{Maximizar} \qquad x_{m+1},$$

sujeta a

$$\begin{aligned}
p_{11}x_1 + p_{21}x_2 + \cdots + p_{m1}x_m - x_{m+1} &\geq 0 \\
p_{12}x_1 + p_{22}x_2 + \cdots + p_{m2}x_m - x_{m+1} &\geq 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots &\cdots \\
p_{1n}x_1 + p_{2n}x_2 + \cdots + p_{mn}x_m - x_{m+1} &\geq 0 \\
x_1 + x_2 + \cdots + x_m &= 1
\end{aligned}$$

y

$$x_i \geq 0, \qquad \text{para } i = 1, 2, \ldots, m.$$

Observe que x_{m+1} no está restringida a ser no negativa, mientras que el método símplex sólo se puede aplicar una vez que *todas* las variables tienen la restricción de no negatividad. Este asunto puede resolverse con facilidad como se verá en seguida.

Ahora considere al jugador 2. Éste puede encontrar su estrategia óptima mixta si reescribe la matriz de pagos como los pagos a sí mismo en lugar de al jugador 1 y procediendo exactamente como se acaba de describir. Sin embargo, resulta ilustrativo resumir su formulación en términos de la matriz de pagos original. Si se sigue un procedimiento análogo al descrito, el jugador 2 concluiría que su estrategia mixta óptima está dada por la solución óptima del problema de programación lineal:

Minimizar $\quad y_{n+1},$

sujeta a

$$
\begin{aligned}
p_{11}y_1 + p_{12}y_2 + \cdots + p_{1n}y_n - y_{n+1} &\leq 0\\
p_{21}y_1 + p_{22}y_2 + \cdots + p_{2n}y_n - y_{n+1} &\leq 0\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots&\cdots\\
p_{m1}y_1 + p_{m2}y_2 + \cdots + p_{mn}y_n - y_{n+1} &\leq 0\\
y_1 + y_2 + \cdots + y_n &= 1
\end{aligned}
$$

y

$$y_j \geq 0, \qquad \text{para } j = 1, 2, \ldots, n.$$

Es sencillo demostrar (vea el problema 14.5-6 y su sugerencia) que este problema de programación lineal y el dado para el jugador 1 son *duales* uno del otro en el sentido descrito en las secciones 6.1 y 6.4. Este hecho tiene varias implicaciones importantes. Una es que se pueden encontrar las estrategias mixtas óptimas para los dos jugadores mediante la resolución de sólo uno de los problemas de programación lineal puesto que la solución óptima dual es un producto complementario automático de los cálculos del método símplex para encontrar la solución óptima primal. Una segunda implicación es que esto trae consigo toda la *teoría de dualidad* (descrita en el capítulo 6) para fundamentar en ella la interpretación y el análisis de los juegos.

Una implicación relacionada es que proporciona una **prueba muy sencilla del teorema minimax**. Sean x^*_{m+1} y y^*_{n+1} los valores de x_{m+1} y y_{n+1} en la solución óptima de los respectivos problemas de programación lineal. Se sabe, por la *propiedad de dualidad fuerte* que se presentó en la sección 6.1, que $-x^*_{m+1} = -y^*_{n+1}$, de manera que $x^*_{m+1} = y^*_{n+1}$. Sin embargo, es evidente de la definición de $\underline{v}$ y $\overline{v}$ que $\underline{v} = x^*_{m+1}$ y que $\overline{v} = y^*_{n+1}$, lo que conduce a la conclusión de que $\underline{v} = \overline{v}$, como lo establece el teorema minimax.

Queda un cabo suelto por atar, esto es, cómo proceder si x_{m+1} y y_{n+1} no están restringidas en signo en sus formulaciones de programación lineal. Si es evidente que $v \geq 0$ para que los valores óptimos de x_{m+1} y y_{n+1} sean no negativos, entonces no hay peligro si se introducen las restricciones de no negatividad sobre estas variables con el propósito de aplicar el método símplex. No obstante, si $v < 0$, entonces debe hacerse un ajuste. Una posibilidad es emplear el enfoque descrito en la sección 4.6 en el que se sustituye una variable no restringida por la diferencia de dos variables no negativas. Otra posibilidad es invertir a los jugadores 1 y 2 para que la matriz de pagos se reescriba como el pago al jugador 2 original, lo que haría que el valor correspondiente de v fuera positivo. Un tercer procedimiento, y el que más se usa en la práctica, es agregar una constante fija grande a todos los elementos de la matriz de pagos para que el nuevo valor del juego sea positivo. (Por ejemplo, bastaría con igualar esta constante al valor absoluto del elemento más negativo.) Como la misma constante se agrega a todos los elementos, este ajuste no puede alterar de ninguna manera las estrategias mixtas óptimas, por lo cual ahora se pueden obtener en forma normal. El valor indicado del juego se aumentará en la cantidad constante, pero se puede reajustar después de obtener la solución.

Aplicación a la variación 3 del problema de las campañas políticas

Para ilustrar este enfoque de programación lineal, considere de nuevo la variación 3 del problema de la campaña política después de eliminar la estrategia dominada 3 del jugador 1 (vea la tabla 14.6). Como existen algunos elementos negativos en la matriz de pagos reducida, no es evidente si el *valor* del juego v es *no negativo* (ocurre que sí lo es). Por el momento, suponga que $v \geq 0$ y proceda sin hacer ninguno de los ajustes mencionados.

Para escribir el modelo de programación lineal del jugador 1 de este ejemplo, observe que p_{ij} en el modelo general es el elemento del renglón i y la columna j de la tabla 14.6, para $i = 1, 2$ y $j = 1, 2, 3$. El modelo que se obtiene es

Maximizar $\quad x_3,$

sujeta a

$$\begin{aligned} 5x_2 - x_3 &\geq 0 \\ -2x_1 + 4x_2 - x_3 &\geq 0 \\ 2x_1 - 3x_2 - x_3 &\geq 0 \\ x_1 + x_2 \quad &= 1 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

La aplicación del método símplex a este problema de programación lineal (después de agregar la restricción $x_3 \geq 0$) da $x_1^* = \frac{7}{11}, x_2^* = \frac{4}{11}, x_3^* = \frac{2}{11}$ como solución óptima. (Vea los problemas 14.5-8 y 14.5-9.) En consecuencia, la estrategia mixta óptima del jugador 1 de acuerdo al criterio minimax es $(x_1, x_2) = (\frac{7}{11}, \frac{4}{11})$ y el valor del juego es $v = x_3^* = \frac{2}{11}$. El método símplex también produce la solución óptima del dual (que se proporciona en seguida) de este problema, ésta es, $y_1^* = 0, y_2^* = \frac{5}{11}, y_3^* = \frac{6}{11}, y_4^* = \frac{2}{11}$, con lo que la estrategia mixta óptima del jugador 2 es $(y_1, y_2, y_3) = (0, \frac{5}{11}, \frac{6}{11})$.

El dual del problema anterior es el modelo de programación lineal del jugador 2 (el que tiene las variables $y_1, y_2, \ldots, y_n, y_{n+1}$) que se mostró antes en esta misma sección. (Vea el problema 14.5-7.) Al sustituir los valores de p_{ij}, dados en la tabla 14.6, este modelo es

Minimizar $\quad y_4,$

sujeta a

$$\begin{aligned} - 2y_2 + 2y_3 - y_4 &\leq 0 \\ 5y_1 + 4y_2 - 3y_3 - y_4 &\leq 0 \\ y_1 + y_2 + y_3 \quad &= 1 \end{aligned}$$

y

$$y_1 \geq 0, \qquad y_2 \geq 0, \qquad y_3 \geq 0.$$

Si se aplica el método símplex directamente a este modelo (después de agregar la restricción $y_4 \geq 0$), se obtiene la solución óptima $y_1^* = 0, y_2^* = \frac{5}{11}, y_3^* = \frac{6}{11}, y_4^* = \frac{2}{11}$ (al igual que la solución óptima dual, $x_1^* = \frac{7}{11}, x_2^* = \frac{4}{11}, x_3^* = \frac{2}{11}$). En consecuencia, la estrategia mixta óptima del jugador 2 es $(y_1, y_2, y_3) = (0, \frac{5}{11}, \frac{6}{11})$ y de nuevo se puede ver que el valor del juego es $v = y_4^* = \frac{2}{11}$.

Como ya se había determinado la estrategia mixta óptima del jugador 2 cuando se resolvió el primer modelo, no fue necesario resolver el segundo. En general, siempre se pueden encontrar las estrategias mixtas óptimas de *ambos* jugadores con sólo elegir uno de los modelos (cualquiera) y usar el método símplex para obtener una solución óptima y una solución óptima dual.

Cuando se aplicó el método símplex a estos dos modelos de programación lineal, se agregó una restricción de no negatividad que suponía que $v \geq 0$. Si este supuesto se violara, ninguno de los dos modelos tendría soluciones factibles, y el método símplex se detendría con rapidez con este mensaje. Para evitar este riesgo se pudo haber agregado una constante positiva, como 3 (el valor absoluto del elemento más negativo), a todos los elementos de la tabla 14.6. Esto habría aumentado en 3 todos los coeficientes de x_1, x_2, y_1, y_2 y y_3 en las restricciones de desigualdad de los dos modelos. (Vea el problema 14.5-2.)

14.6 EXTENSIONES

Aunque en este capítulo se han considerado sólo los juegos de dos personas y suma cero con un número finito de estrategias puras, la teoría de juegos no se limita a este tipo de juegos. En realidad, se han llevado a cabo investigaciones extensas sobre varios tipos de juegos más complicados, que incluyen los que se resumen en esta sección.

La generalización más sencilla es el *juego de dos personas* con *suma constante*. En este caso, la suma de los pagos de los dos jugadores es una constante fija (positiva o negativa) sin que importe qué combinación de estrategias se seleccione. La única diferencia con un juego de dos personas con suma cero es que, en este caso, la constante debe ser cero. En su lugar, puede surgir una constante diferente de cero debido a que, además de que un jugador gane lo que el otro pierde, los dos jugadores pueden compartir alguna recompensa (si la constante es positiva) o algún costo (si la constante es negativa) por participar en el juego. Agregar esta constante fija no afecta la estrategia que debe elegirse. Por lo tanto, el análisis para determinar las estrategias óptimas es idéntico al descrito en este capítulo para los juegos de dos personas y suma cero.

Una extensión más complicada es el *juego de n-personas*, en el que pueden participar más de dos jugadores. Esta generalización es de particular importancia puesto que, con frecuencia, en muchas situaciones competitivas, se encuentran involucrados más de dos competidores. Esto puede ocurrir, por ejemplo, en la competencia entre empresas de negocios, en la diplomacia internacional, etc. Desafortunadamente, la teoría existente para tales juegos es mucho menos satisfactoria que la de juegos de dos personas.

Otra generalización es el *juego de suma no cero*, en el que la suma de los pagos a los jugadores no tiene que ser 0 (o ninguna otra constante fija). Este caso refleja el hecho de que muchas situaciones de competencia incluyen aspectos no competitivos que contribuyen a la ventaja o desventaja mutua de los jugadores. Por ejemplo, las estrategias de publicidad de compañías que compiten por un mismo mercado pueden afectar no sólo la distribución de ese mercado sino también el tamaño total del mercado que comparte sus productos. Sin embargo, al contrario de un juego de suma constante, el tamaño de la ganancia mutua (o pérdida) para los jugadores depende de la combinación de las estrategias elegidas.

Como es posible la ganancia mutua, los juegos de suma no cero se subdividen en términos del grado en que se permite que los jugadores cooperen. En un extremo se encuentra el *juego no cooperativo*, donde no hay comunicación entre los jugadores anterior al juego. En el otro extremo está el *juego cooperativo* en el que se permite análisis y acuerdos antes del juego. Ejemplos que se podrían formular como juegos cooperativos son las situaciones competitivas que engloban leyes de comercio exterior entre países o los acuerdos que se toman entre empresas y obreros. Cuando existen más de dos jugadores, los juegos cooperativos permiten también que se formen coaliciones entre algunos o todos los jugadores.

Otra extensión más es la que se llama de *juegos infinitos*, donde los jugadores cuentan con un número infinito de estrategias puras. Estos juegos están diseñados para aquellas situaciones en las que la estrategia seleccionada se puede representar mediante una variable de decisión *continua*. Por ejemplo, la variable de decisión puede ser el tiempo en el que se lleva a cabo cierta acción, o la proporción de recursos propios que se asignan a cierta actividad, en una situación de competencia.

Sin embargo, el análisis que se requiere en estas extensiones que están más allá del juego finito de dos personas con suma cero, es relativamente complejo y no se profundizará en él. (Para mayor información al respecto, consulte las referencias seleccionadas 4, 6, 7, 8 y 10.)

14.7 CONCLUSIONES

El problema general de cómo tomar una decisión en un medio competitivo es bastante común e importante. La contribución fundamental de la teoría de juegos es que proporciona un marco conceptual básico para formular y analizar tales problemas en situaciones sencillas. Sin embargo, existe un gran abismo entre lo que la teoría puede manejar y la complejidad de la mayor parte de las situaciones de competencia que surgen en la práctica. Ello significa que, por lo general, las herramientas conceptuales de la teoría de juegos desempeñan un papel complementario cuando se aplican a esas situaciones.

Dada la importancia general del problema, la investigación sobre este tema continúa con algunos éxitos y pretende extender la teoría a casos más complejos.

REFERENCIAS SELECCIONADAS

1. Aumann, R. J. y S. Hart (eds.): *Handbook of Game Theory: With Application to Economics*, vols. 1, 2 y 3, North-Holland, Ámsterdam, 1992, 1994, 2002.

2. Brandenburger, A. y H. Stuart: “Biform Games”, en *Management Science*, **53**(4): 537-549, abril de 2007.
3. Chatterjee, K. y W. F. Samuelson (eds.): *Game Theory and Business Applications*, Kluwer Academic Publishers, Boston, 2001.
4. Forgó, F., J. Szép y F. Szidarovsky: *Introduction to the Theory of Games: Concepts, Methods, Applications*, Kluwer Academic Publishers, Boston, 1999.
5. Hohzaki, R.: “A Search Game Taking Account of Attributes of Searching Resources”, en *Naval Research Logistics*, **55**(1): 76-90, febrero de 2008.
6. Mendelson, E.: *Introducing Game Theory and Its Applications*, Chapman and Hall/CRC Press, Boca Raton, FL, 2005.
7. Meyerson, R. B.: *Game Theory: Analysis of Conflict*, Harvard University Press, Cambridge, MA, 1991.
8. Owen, G.: *Game Theory*, 3a. ed., Academic Press, San Diego, 1995.
9. Parthasarathy, T., B. Dutta y A. Sen (eds.): *Game Theoretical Applications to Economics and Operations Research*, Kluwer Academic Publishers, Boston, 1997.
10. Webb, J. N.: *Game Theory: Decisions, Interaction and Evolution*, Springer, Nueva York, 2007.
11. Zhuang, J., y V. M. Bier: “Balancing Terrorism and Natural Disasters-Defensive Strategy with Endogenous Attacker Effort”, en *Operations Research*, **55**(5): 976-991, septiembre-octubre de 2007.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos para el capítulo 14.

Archivos para resolver los ejemplos “Ch. 14—Game Theory”:

Archivos de Excel
Archivos de LINGO/LINDO
Archivos de MPL/CPLEX

Glosario del capítulo 14

Vea en el apéndice 1 la documentación del software.

PROBLEMAS

Los símbolos a la izquierda de algunos problemas (o de sus incisos) significan lo siguiente:

C: Use la computadora con cualquier opción de software disponible (o la recomendada por el profesor) para resolver el problema.

Un asterisco en el número del problema indica que al final del libro se da al menos una respuesta parcial.

14.1-1. El sindicato y la administración de una compañía negocian el nuevo contrato colectivo. Por ahora las negociaciones están congeladas, pues la empresa ha hecho una oferta “final” de un aumento salarial de $1.10 por hora y el sindicato una demanda “final” de un aumento de $1.60 por hora. Ambas partes han acordado que un árbitro imparcial establezca el aumento en alguna cantidad entre $1.10 por hora y $1.60 por hora (inclusive).

El arbitraje ha pedido a cada parte que le presente una propuesta confidencial de un aumento salarial económicamente razonable y justo (redondeado a los diez centavos más cercanos). Por experiencias anteriores, ambas partes saben que por lo general el árbitro acepta la propuesta del lado que cede más en su cifra final. Si ningún lado cambia su cantidad final o si ambos ceden en la misma cantidad, el arbitraje suele establecer una cifra a la mitad ($1.35 en este caso). Ahora, cada parte necesita determinar qué aumento proponer para obtener un beneficio máximo.

Formule este problema como un juego de dos personas y suma cero.

14.1-2. Dos fabricantes compiten por las ventas de dos líneas de productos distintas pero igualmente redituables. En ambos casos, el volumen de ventas del fabricante 2 es el triple del que logra el fabricante 1. En vista de algunos avances tecnológicos, ambos harán mejoras importantes a los dos productos, pero no están seguros de la estrategia de desarrollo y comercialización que deben seguir.

Si desarrollan al mismo tiempo las mejoras de los dos productos, ningún fabricante podrá tenerlos listos para la venta antes de 12 meses. Una alternativa es llevar a cabo un “programa intensivo” para desarrollar primero uno de los dos productos y tratar de comercializarlo antes de que la competencia lo haga. Si actúa de

esta manera, el fabricante 2 podría tener un producto listo para la venta en nueve meses, mientras que el fabricante 1 requeriría 10 meses (por compromisos previos de sus instalaciones). Cualquiera de los fabricantes podría tener el segundo producto listo en otros nueve meses.

Para cualquier línea de producto, si los dos fabricantes comercializan los modelos mejorados simultáneamente, se estima que el fabricante 1 aumentaría 8% (de 25 a 33%) el porcentaje del total de las ventas futuras de este producto. De la misma manera, el fabricante 1 aumentaría sus ventas 20, 30 y 40% del total si comercializa el producto 2, seis y ocho meses antes que el fabricante 2, respectivamente. Por otro lado, el fabricante 1 perdería 4, 10, 12 y 14% del total si el fabricante 2 logra comercializar uno, tres, siete y diez meses antes que él.

Formule este problema como un juego de suma cero de dos personas y determine la estrategia que deben seguir los dos fabricantes según el criterio minimax.

14.1-3. Considere un juego de mesa entre dos personas. Cada una comienza con tres fichas: una roja, una blanca y una azul. Cada ficha se puede usar una sola vez.

Para comenzar, cada jugador elige una de sus fichas y la coloca sobre la mesa, tapada; después ambos la destapan y determinan el pago para el ganador. En particular, si ambos tienen el mismo color, es un empate; de otra manera, la tabla que sigue indica el ganador y el pago que debe recibir del otro jugador. En seguida, cada jugador elige una de sus dos fichas restantes y se repite el proceso con un nuevo pago de acuerdo con la tabla. Por último, cada jugador juega la ficha que le queda y se determina el tercer pago, que es el final.

Ficha ganadora	Pago ($)
Rojo gana a blanco	90
Blanco gana a azul	70
Azul gana a rojo	50
Fichas iguales	0

Formule este problema como un juego de suma cero entre dos personas, e identifique la forma de las estrategias y pagos.

14.2-1. Reconsidere el problema 14.1-1.

a) Utilice el concepto de estrategias dominadas para determinar la mejor estrategia de cada parte.

b) Sin eliminar las estrategias dominadas, use el criterio minimax para determinar la mejor estrategia de cada jugador.

14.2-2.* Para la siguiente matriz de pagos determine la estrategia óptima de cada jugador. Para ello elimine de manera sucesiva las estrategias dominadas. (Indique el orden en el que elimina las estrategias.)

		Jugador 2		
Estrategia		1	2	3
	1	−3	1	2
Jugador 1	2	1	2	1
	3	1	0	−2

14.2-3. Considere el juego que tiene la siguiente matriz de pagos.

		Jugador 2			
Estrategia		1	2	3	4
	1	2	−3	−1	1
Jugador 1	2	−1	1	−2	2
	3	−1	2	−1	3

Determine la estrategia óptima de cada jugador. Para ello elimine de manera sucesiva las estrategias dominadas. Proporcione una lista de estas estrategias dominadas (y la estrategia dominante correspondiente) en el orden en el que las eliminó.

14.2-4. Encuentre el punto silla del juego que tiene la siguiente matriz de pagos.

		Jugador 2		
Estrategia		1	2	3
	1	1	−1	1
Jugador 1	2	−2	0	3
	3	3	1	2

Utilice el criterio minimax para encontrar la mejor estrategia de cada jugador. ¿Tiene este juego un punto silla? ¿Se trata de un juego estable?

14.2-5. Encuentre el punto silla del juego que tiene la siguiente matriz de pagos.

		Jugador 2			
Estrategia		1	2	3	4
	1	3	−3	−2	−4
Jugador 1	2	−4	−2	−1	1
	3	1	−1	2	0

Aplique el criterio minimax para encontrar la mejor estrategia de cada jugador. ¿Tiene este juego un punto silla? ¿Se trata de un juego estable?

14.2-6. Dos compañías comparten el grueso del mercado de cierto tipo de producto. Cada una elabora nuevos planes de comercialización para el próximo año con la intención de arrebatar parte de las ventas a la otra compañía. (Las ventas totales del producto son más o menos fijas, por lo que una compañía puede incrementar sus ventas sólo si disminuyen las de la otra.) Cada una considera tres posibilidades: 1) un mejor empaque del producto, 2) un aumento de publicidad y 3) una pequeña reducción de precio. Los costos de las tres opciones son comparables y lo suficientemente grandes como para que cada compañía elija sólo una. El efecto estimado de cada combinación de alternativas sobre el *aumento del porcentaje de las ventas* de la compañía 1 es:

		Jugador 2		
Estrategia		**1**	**2**	**3**
	1	2	3	1
Jugador 1	2	1	4	0
	3	3	−2	−1

Cada compañía debe hacer su elección antes de conocer la decisión de la otra compañía.

a) Sin eliminar las estrategias dominantes utilice el criterio minimax (o maximin) para determinar la mejor estrategia de cada compañía.

b) Ahora identifique y elimine las estrategias dominadas hasta donde sea posible. Elabore una lista de las estrategias dominadas que muestre el orden en el que se pudieron eliminar. Después elabore la matriz de pagos reducida que resulta cuando ya no quedan estrategias dominantes.

14.2-7.* Dos políticos están a punto de comenzar sus campañas y compiten por el mismo puesto. Cada uno debe elegir el aspecto principal sobre el que hará hincapié en el programa de trabajo. Cada uno tiene tres temas ventajosos entre los que puede elegir, pero su eficacia relativa dependerá del que seleccione su oponente. En particular, el aumento estimado del número de votos para el político 1 (expresado en porcentaje de la vocación total) resultado de cada combinación de temas es el siguiente:

		Tema político 2		
		1	**2**	**3**
Tema político 1	1	7	−1	3
	2	1	0	2
	3	−5	−3	−1

No obstante, como se requieren muchas horas-hombre para investigar y formular el tema elegido, cada político deberá hacer su propia elección antes de conocer la de su oponente. ¿Qué tema debe elegir?

Para cada una de las situaciones descritas a continuación, formule este problema como un juego de suma cero entre dos personas; después determine el tema que debe elegir cada político de acuerdo con el criterio específico.

a) Existe gran incertidumbre respecto de las preferencias actuales de los votantes, por lo que cada punto porcentual adicional de votos ganados por un político tiene el mismo valor para él. Utilice el criterio minimax.

b) Una encuesta confiable determinó que el porcentaje de votantes que en este momento prefieren al político 1 (antes de presentar sus programas de trabajo) se encuentra entre 45 y 50%. (Suponga una distribución uniforme en este intervalo.) Utilice el concepto de estrategias dominadas; comience con las estrategias del político 1.

c) Suponga que el porcentaje descrito en *b*) fuera en realidad de 45%. ¿Debe el político 1 usar el criterio minimax? Dé una explicación. ¿Qué tema se le puede recomendar? ¿Por qué?

14.2-8. Describa brevemente lo que se piensa sobre las ventajas y desventajas del criterio minimax.

14.3-1. Considere el juego de pares e impares que se presentó en la sección 14.1 y cuya tabla de pagos se muestra en la figura 14.1.

a) Demuestre que este juego no tiene punto de silla.

b) Escriba una expresión del pago esperado para el jugador 1 (el jugador par) en términos de las probabilidades de que los dos jugadores usen sus estrategias puras respectivas. Después, muestre a qué se reduce esta expresión en los tres casos siguientes: *i*) el jugador 2 definitivamente utiliza su primera estrategia, *ii*) el jugador 2 definitivamente utiliza su segunda estrategia, *iii*) el jugador 2 asigna probabilidades iguales a usar sus dos estrategias.

c) Repita la parte *b*) cuando el jugador 1 se convierte en el jugador impar.

14.3-2. Considere el siguiente juego de mesa entre dos jugadores. Un árbitro tira una moneda al aire, anota si cae cara o cruz y la muestra sólo al jugador 1. Este jugador puede: *i*) pasar y pagar 5 dólares al jugador 2 o *ii*) apostar. Si el jugador 1 pasa, el juego se termina, pero si apuesta, el juego sigue y el jugador 2 puede: *i*) pasar y pagar 5 dólares al jugador 1, o *ii*) apostar. Si el jugador 2 apuesta, el árbitro le muestra la moneda; si es cara, el jugador 2 paga 10 dólares al jugador 1; si es cruz, el jugador 2 recibe 10 dólares del jugador 1.

a) Proporcione las estrategias puras de cada jugador. (*Sugerencia:* El jugador 1 tendrá cuatro estrategias puras que especifican cómo respondería a cada uno de los dos resultados que le muestra el árbitro; el jugador 2 tendrá dos estrategias puras que especifican cómo respondería si el jugador 1 apuesta.)

b) Desarrolle la matriz de pagos de este juego con valores esperados de los elementos, cuando sea necesario. Identifique y elimine las estrategias dominadas.

c) Muestre que ninguno de los elementos de la matriz de pagos que se obtuvo son punto silla. Explique por qué cualquier elección fija de una estrategia pura de cada jugador tiene que ser una solución inestable, por lo que, en su lugar, deben usarse estrategias mixtas.

d) Escriba una expresión para el pago esperado en términos de las probabilidades de los dos jugadores con sus estrategias respectivas. Muestre a qué se reduce esta expresión en los siguientes casos: *i*) el jugador 2 usa su primera estrategia, *ii*) el jugador 2 usa su segunda estrategia, *iii*) el jugador 2 asigna probabilidades iguales al empleo de sus dos estrategias.

14.4-1. Considere el juego de pares e impares que se presentó en la sección 14.1 y cuya tabla de pagos se muestra en la tabla 14.1. Use el procedimiento gráfico que se describió en la sección 14.4 desde el punto de vista del jugador 1 (el jugador par) para determinar la estrategia mezclada óptima de cada jugador de acuerdo con el criterio minimax. Después, repita lo anterior desde el punto de vista del jugador 2 (el jugador impar). También proporcione el valor correspondiente del juego.

14.4-2. Reconsidere el problema 14.3-2. Utilice el procedimiento gráfico que se describe en la sección 14.4 para determinar la estrategia mixta óptima de cada jugador según el criterio minimax. Además, proporcione el valor correspondiente del juego.

14.4-3. Considere el juego que tiene la siguiente matriz de pagos.

		Jugador 2	
Estrategia		**1**	**2**
Jugador 1	1	5	−4
	2	−2	3

Utilice el procedimiento gráfico descrito en la sección 14.4 para determinar el valor del juego y la estrategia mixta óptima de cada jugador según el criterio minimax. Verifique la respuesta del jugador 2 con su matriz de pagos y la aplicación del procedimiento gráfico a esta tabla, de manera directa.

14.4-4.* Para la siguiente matriz de pagos utilice el procedimiento gráfico descrito en la sección 14.4 para determinar el valor del juego y la estrategia mixta óptima de cada jugador según el criterio minimax.

		Jugador 2		
Estrategia		**1**	**2**	**3**
Jugador 1	1	4	3	1
	2	0	1	2

14.4-5. El equipo A. J. Swim está en vísperas de una importante competencia de natación contra el equipo G. N. Swim. Cada equipo cuenta con un nadador estrella (John y Mark, respectivamente), que tienen un excelente desempeño en las pruebas de 100 m mariposa, dorso y pecho. Sin embargo, las reglas no permiten que tomen parte en más de dos eventos. Sus entrenadores necesitan decidir cómo usarlos para obtener la mayor ventaja.

Cada equipo inscribirá a tres nadadores por prueba (el máximo permitido). La siguiente tabla proporciona los mejores tiempos logrados por John y Mark en cada prueba, junto con los mejores tiempos de otros nadadores que participarán en el evento. (En la prueba en la que John o Mark no sean incluidos, el tercer elemento de su equipo será más lento que los dos que se dan en la tabla.)

	Equipo A. J. Swim			Equipo G. N. Swim		
	Carrera			Carrera		
	1	**2**	**John**	**Mark**	**1**	**2**
Mariposa	1:01.6	59.1	57.5	58.4	1:03.2	59.8
Dorso	1:06.8	1:05.6	1:03.3	1:02.6	1:04.9	1:04.1
Pecho	1:13.9	1:12.5	1:04.7	1:06.1	1:15.3	1:11.8

Se obtienen 5 puntos por el primer lugar, 3 por el segundo, 1 por el tercero y ningún punto por otros lugares. Ambos entrenadores piensan que, en esencia, todos los nadadores igualarán sus mejores marcas en este encuentro. Así, John y Mark quedarán, definitivamente, en dos de estos tres eventos.

a) Los entrenadores deben presentar los nombres de sus competidores antes del encuentro sin saber los del otro equipo y no se permiten cambios. El resultado del encuentro es incierto, por lo que, para los entrenadores, cada punto adicional tiene el mismo valor. Formule este problema como un juego de suma cero entre dos personas. Elimine las estrategias dominantes y después utilice el procedimiento gráfico descrito en la sección 14.4 para encontrar las estrategias mixtas óptimas de cada equipo, de acuerdo con el criterio minimax.

b) La situación y asignación son iguales que en el inciso *a*), excepto que ambos entrenadores piensan que el equipo de A. J. Swim triunfará en el encuentro si puede ganar 13 puntos o más en los tres eventos, pero perderá con menos de 13 puntos. [Compare las estrategias mixtas óptimas con las que obtuvo en el inciso *a*).]

c) Ahora suponga que los entrenadores pueden presentar los nombres de sus competidores durante el encuentro, prueba por prueba. Cuando entregan los nombres de los competidores de una prueba no saben quién estará incluido por parte del otro equipo, pero sí saben quién nadó en las pruebas *anteriores*. Las tres pruebas se llevan a cabo en el orden en que se dan en la tabla. Una vez más, el equipo de A. J. Swim necesita 13 puntos en estos eventos para ganar el encuentro. Formule el problema como un juego entre dos personas y suma cero. Use el concepto de estrategias dominadas para determinar la mejor estrategia del equipo G. N. que de hecho "garantice" que ganará bajo los supuestos establecidos.

d) La situación es la misma que en el inciso *c*), pero ahora suponga que el entrenador del equipo G. N. no sabe nada de teoría de juegos y puede elegir cualquiera de sus estrategias en las que Mark nada en dos pruebas. Use el concepto de estrategias dominadas para determinar las mejores estrategias entre las que deba elegir el equipo A. J. Si este entrenador sabe que el otro tiende a colocar a Mark en mariposa con más frecuencia que en dorso, ¿qué estrategia debe elegir?

14.5-1. Considere el juego de pares e impares que se presentó en la sección 14.1 y cuya tabla de pagos se muestra en la tabla 14.1.

a) Use el método que se describió en la sección 14.5 para formular el problema consistente en encontrar las estrategias mezcladas óptimas de acuerdo con el criterio mínimax como dos problemas de programación lineal, uno para el jugador 1 (el jugador par) y el otro para el jugador 2 (el jugador impar) como el dual del primer problema.

b) Utilice el método símplex para determinar la mezcla óptima de estrategias.

14.5-2. En referencia al último párrafo de la sección 14.5, suponga que se sumó 3 a todos los elementos de la tabla 14.6 con el fin de asegurar que los modelos de programación lineal de ambos jugadores tuvieran soluciones factibles con $x_3 \geq 0$ y $y_4 \geq 0$. Escriba los dos modelos. Con base en la información que se proporcionó en la sección 14.5, ¿cuáles son las soluciones óptimas de estos dos modelos? ¿Cuál es la relación entre x_3^* y y_4^*? ¿Cuál es la relación entre el *valor* del juego original v y los valores de x_3^* y y_4^*?

14.5-3.* Considere el juego que tiene la siguiente matriz de pagos.

		Jugador 2			
Estrategia		**1**	**2**	**3**	**4**
	1	5	0	3	1
Jugador 1	2	2	4	3	2
	3	3	2	0	4

a) Utilice el enfoque descrito en la sección 14.5 para formular el problema de encontrar las estrategias mixtas óptimas de acuerdo con el criterio minimax, como un problema de programación lineal.

C ***b***) Aplique el método símplex para encontrar estas estrategias mixtas.

14.5-4. Siga las instrucciones del problema 14.5-3 para el juego que tiene la siguiente matriz de pagos.

Estrategia		Jugador 2: 1	2	3
	1	4	2	−3
Jugador 1	2	−1	0	3
	3	2	3	−2

14.5-5. Siga las instrucciones del problema 14.5-3 para el juego que tiene la siguiente matriz de pagos.

Estrategia		Jugador 2: 1	2	3	4	5
	1	1	−3	2	−2	1
Jugador 1	2	2	3	0	3	−2
	3	0	4	−1	−3	2
	4	−4	0	−2	2	−1

14.5-6. En la sección 14.5 se presentó una formulación de programación lineal para encontrar la estrategia mixta óptima de los jugadores 1 y 2. Use la tabla 6.14 y demuestre que el problema de programación lineal del jugador 2 es el dual del problema del jugador 1. (*Sugerencia:* Recuerde que una variable dual con una restricción de no positividad $y_i' \le 0$ se puede sustituir por $y_i = -y_i'$ con una restricción de no negatividad $y_i \ge 0$.)

14.5-7. Considere los modelos de programación lineal de los jugadores 1 y 2 que se dio casi al final de la sección 14.5 en la variación 3 del problema de la campaña política (vea la tabla 14.6). Siga las instrucciones del problema 14.5-6 para manejar estos dos modelos.

14.5-8. Considere la variación 3 del problema de la campaña política (tabla 14.6). Haga referencia al problema de programación lineal del jugador 1 que se obtuvo casi al final de la sección 14.5. Pase por alto la variable de la función objetivo x_3, grafique la *región factible* de x_1 y x_2 (como se describió en la sección 3.1.) (*Sugerencia:* Esta región factible es un segmento de recta.) Escriba una expresión algebraica para el valor de x_3 que maximiza, para cualquier punto en esta región factible. Por último, use esta expresión para demostrar que la solución óptima debe, en realidad, ser la que se obtuvo en la sección 14.5.

C **14.5-9.** Considere el modelo de programación lineal del jugador 1 que se proporcionó al final de la sección 14.5 en la variación 3 del problema de la campaña política (tabla 14.6). Verifique las estrategias mixtas óptimas de ambos jugadores que se dieron en la sección 14.5 mediante la aplicación de una rutina automática del método símplex a este modelo para encontrar la solución óptima y la solución óptima dual.

14.5-10. Considere el juego general $m \times n$ de dos personas con suma cero. Sea p_{ij} el pago al jugador 1 si juega la estrategia i ($i = 1, \ldots, m$) y el jugador 2 juega su estrategia j ($j = 1, \ldots, n$). Se dice que, por ejemplo, la estrategia 1 del jugador 1 es dominada débilmente por la estrategia 2 si $p_{1j} \le p_{2j}$ para $j = 1, \ldots, n$ y $p_{1j} = p_{2j}$ para uno o más valores de j.

a) Suponga que la matriz de pagos tiene uno o más puntos silla, de manera que los jugadores tienen estrategias puras óptimas correspondientes, de acuerdo con el criterio minimax. Demuestre que si se eliminan las estrategias *dominadas* débilmente de la matriz de pagos, no se eliminan todos estos puntos silla y no pueden producirse nuevos.

b) Suponga que la matriz de pagos no tiene puntos silla, de manera que las estrategias óptimas con el criterio minimax son estrategias mixtas. Demuestre que si se eliminan las estrategias puras dominadas débilmente de la matriz de pagos, no se eliminan todas las estrategias mixtas óptimas y no se pueden producir nuevas.

CAPÍTULO 15

Análisis de decisiones

Los capítulos anteriores se han centrado en la toma de decisiones cuando las consecuencias de las decisiones alternativas se conocen con un grado razonable de certidumbre. Este entorno de toma de decisiones permitió formular modelos matemáticos útiles —programación lineal, programación entera, programación no lineal, etc.— con funciones objetivo que especifican las consecuencias estimadas de cualquier combinación de decisiones. Aunque estas consecuencias casi nunca pueden predecirse con total seguridad, al menos se pueden estimar con suficiente exactitud para justificar el uso de estos modelos (junto con el análisis de sensibilidad, etc.).

Sin embargo, muchas veces las decisiones deben tomarse en entornos con mayor incertidumbre. Se enumeran algunos ejemplos.

1. Un fabricante introduce un nuevo producto al mercado. ¿Cuál será la probable reacción de los consumidores? ¿Cuánto debe producir? ¿Debe probar el producto en una región pequeña antes de decidir la distribución integral? ¿Cuánta publicidad necesita para lanzar el producto con éxito?
2. Una empresa financiera invierte en certificados. ¿Cuáles son los mejores prospectos de certificados de sectores del mercado e individuales? ¿Hacia dónde va la economía? ¿Cuáles son las tasas de interés? ¿Cómo afectan estos factores las decisiones de inversión?
3. Un contratista del gobierno se presenta a una licitación. ¿Cuáles serán los costos reales del proyecto? ¿Qué otras compañías se han presentado? ¿Cuál es su presupuesto probable?
4. Una empresa agrícola selecciona la mezcla de cosechas y ganado para la próxima temporada. ¿Cuáles serán las condiciones climáticas? ¿Hacia dónde van los precios? ¿Cuáles serán los costos?
5. Una compañía petrolera decide perforar en cierta región. ¿Cuán probable es que haya petróleo en ella? ¿Cuánto? ¿Cuán profundo tendrá que perforar? ¿Deben los geólogos investigar más el sitio antes de perforar?

El *análisis de decisiones* se diseñó para estudiar estos tipos de decisiones que se deben tomar en un ambiente de gran incertidumbre. Esta herramienta proporciona un marco de trabajo y una metodología para la toma de decisiones racional cuando los resultados son inciertos.

En el capítulo 14 se describe la forma en que la teoría de juegos también se puede usar para enfrentar cierto tipo de toma de decisiones con incertidumbre. Existen algunas similitudes en los enfoques de la teoría de juegos y el análisis de decisiones. No obstante, también existen diferencias porque están diseñados para distintos tipos de aplicaciones. Estas similitudes y diferencias se describirán en la sección 15.2.

Una pregunta que surge con frecuencia es sobre el tiempo para tomar una decisión, si es adecuado en este momento o antes hacer algunas *pruebas* (con cierto costo) para reducir el nivel de incertidumbre sobre el resultado de la decisión. Por ejemplo, la prueba puede ser una promoción de prueba de un nuevo producto para ver la reacción del consumidor antes de tomar la decisión de proceder o no con la producción y comercialización a gran escala del producto. Se conoce a esta

prueba como *experimentación* y así nos referiremos a ella. En consecuencia, el análisis de decisiones divide la toma de decisiones en los casos *sin experimentación* y *con experimentación*.

En la primera sección se introduce un ejemplo prototipo que se estudiará a través de todo el capítulo con fines ilustrativos. En las secciones 15.2 y 15.3 se presentan los principios básicos de la *toma de decisiones sin experimentación* y de la *toma de decisiones con experimentación*. Después se exponen los *árboles de decisión*, una herramienta útil para describir y analizar el proceso de decisiones cuando debe tomarse una serie de ellas. En la sección 15.5 se explica cómo se realiza el análisis de sensibilidad para árboles de decisión mediante hojas de cálculo. En la sección 15.6 se introduce la *teoría de la utilidad*, que proporciona una manera de calibrar los posibles resultados de la decisión para reflejar el valor verdadero de estos resultados para quien toma las decisiones. El capítulo concluye con la evaluación de la aplicación práctica del análisis de decisiones y un resumen de la variedad de aplicaciones que, en gran medida, han beneficiado a las organizaciones involucradas.

15.1 EJEMPLO PROTOTIPO

La GOFERBROKE COMPANY es dueña de unos terrenos en los que puede haber petróleo. Un geólogo consultor ha informado a la administración que piensa que existe una posibilidad de 1 entre 4 de encontrar petróleo.

Debido a esta posibilidad, otra compañía petrolera ha ofrecido comprar las tierras en 90 000 dólares. Sin embargo, la Goferbroke considera conservarlas para perforar ella misma. El costo de la perforación es de 100 000 dólares. Si encuentra petróleo, el ingreso esperado será de 800 000 dólares; así, la ganancia esperada para la compañía (después de deducir el costo de la perforación) será de 700 000 dólares. Se incurrirá en una pérdida de 100 000 dólares (el costo de barrenar) si no se encuentra petróleo.

En la tabla 15.1 se resumen estos datos. La sección 15.2 estudia cómo abordar la decisión de si se debe perforar o vender con base sólo en estos datos. (Éste será el **primer problema de Goferbroke Co.**)

Sin embargo, antes de decidir entre perforar o vender, otra opción es llevar a cabo una exploración sismológica detallada en el área para obtener una mejor estimación de la probabilidad de encontrar petróleo. (Este proceso de decisión se conoce con el nombre de **problema de Goferbroke completo**.) En la sección 15.3 se presenta este caso de *toma de decisiones con experimentación*, momento en el cual se proporcionarán los datos adicionales necesarios.

Esta compañía opera sin mucho capital por lo que una pérdida de 100 000 dólares sería bastante seria. En la sección 15.6 se describe una manera de perfeccionar la evaluación de las consecuencias de los resultados posibles.

15.2 TOMA DE DECISIONES SIN EXPERIMENTACIÓN

Antes de buscar una solución al problema de la Goferbroke Co., se formulará un marco de referencia general para la toma de decisiones.

En términos generales, el tomador de decisiones debe elegir una **opción** de entre un conjunto de acciones posibles. El conjunto contiene todas las *alternativas factibles* bajo consideración de las distintas formas de proceder en el problema en cuestión.

TABLA 15.1 Ganancias prospectivas para la Goferbroke Company

Alternativa \ Estado del terreno	Pago: Petróleo	Pago: Seco
Perforar buscando petróleo	$700 000	−$100 000
Vender el terreno	$ 90 000	$ 90 000
Posibilidad del estado	1 de 4	3 de 4

Esta elección de una alternativa debe hacerse frente a la incertidumbre porque el resultado se verá afectado por factores aleatorios que se encuentran fuera del control del tomador de decisiones. Estos factores aleatorios determinan qué situación se encontrará en el momento en que se ejecute la acción. Cada una de estas situaciones posibles se conoce como **estado de la naturaleza**.

En el caso de cada combinación de una opción y un estado de la naturaleza, el tomador de decisiones sabe cuál sería el **pago** resultante. El pago es una medida cuantitativa del valor de las consecuencias del resultado para el tomador de decisiones. Por ejemplo, muchas veces el pago se representa por la *ganancia monetaria neta* (utilidad), aunque también se pueden usar otras medidas (como se describe en la sección 15.6). Si las consecuencias del resultado no son por completo ciertas aunque el estado de la naturaleza esté dado, el pago se convierte en un *valor esperado* (en el sentido estadístico) de la medida de las consecuencias. En general, se usa una **tabla de pagos** que indica el pago de cada combinación de opción y estado de la naturaleza.

Si el lector ya estudió teoría de juegos (capítulo 14), debe señalarse una analogía interesante entre estos conceptos de análisis de decisiones y el juego de dos personas con suma cero que se describió en el capítulo 14. El *tomador de decisiones* y la *naturaleza* se pueden ver como *dos jugadores* de este juego. Las *opciones alternativas* y los *estados de la naturaleza* posibles se pueden ver como las *estrategias* disponibles para los respectivos jugadores, donde el resultado de cada combinación de estrategias es un *pago* para el jugador 1 (el tomador de decisiones). Desde este punto de vista, el marco conceptual del análisis de decisiones se puede resumir como:

1. El *tomador de decisiones* necesita elegir una de las *opciones posibles*.
2. Luego, la *naturaleza* elegirá uno de los *estados de la naturaleza* posibles.
3. Cada combinación de una opción y un estado de la naturaleza da como resultado un *pago*, que está dado como uno de los elementos de la *tabla de pagos*.
4. Esta tabla de pagos debe usarse para encontrar una *opción óptima* para el tomador de decisiones según un criterio adecuado.

Más adelante se presentarán tres posibilidades para este criterio, donde el primero (el criterio de pago maximin) se deriva de la teoría de juegos.

Sin embargo, esta analogía con los juegos de dos personas y suma cero pone de manifiesto un aspecto importante. En la teoría de juegos se supone que *ambos* jugadores son *racionales* y eligen sus estrategias para *promover su propio beneficio*. Esta descripción se ajusta al tomador de decisiones pero no a la naturaleza. Por el contrario, la naturaleza es un jugador pasivo que elige sus estrategias (estados de la naturaleza) de alguna manera aleatoria. Este cambio significa que el criterio de la teoría de juegos para la forma de elegir una estrategia óptima (acción) no será la mejor para muchos tomadores de decisiones en el contexto actual.

Es necesario agregar otro elemento a estos conceptos de teoría de decisiones. El tomador de decisiones casi siempre tendrá alguna información que debe tomar en cuenta sobre la posibilidad relativa de los estados de la naturaleza. Es común que se pueda traducir esta información en una distribución de probabilidad, si se piensa que el estado de la naturaleza es una variable aleatoria, en cuyo caso esta distribución se conoce como **distribución *a priori***. Las distribuciones *a priori* con frecuencia son subjetivas pues suelen depender de la experiencia o la intuición de un individuo. Las probabilidades de los respectivos estados de la naturaleza se llaman **probabilidades *a priori***.

Formulación del ejemplo prototipo en este contexto

Como se indica en la tabla 15.1, la Goferbroke Co. debe considerar dos opciones posibles: perforar en busca de petróleo o vender el terreno. Los estados posibles de la naturaleza son que el subsuelo contenga petróleo y que no sea así, lo cual se designa en los encabezados de las columnas de la tabla 15.1 como *petróleo* y *seco*. Debido a que el geólogo consultor ha estimado que existe una oportunidad entre 4 de que haya petróleo (y por ende 3 oportunidades de 4 de que no lo haya), las respectivas probabilidades *a priori* son 0.25 y 0.75. Por lo tanto, con las unidades de pago en miles de dólares de ganancia, la tabla de pagos que se muestra en la tabla 15.1 se puede obtener en forma directa de la tabla 15.2.

Se usará esta tabla de pagos para encontrar la opción óptima de acuerdo con cada uno de los tres criterios que se describirán.

Recuadro de aplicación

A partir de la fusión de Conoco, Inc. y Phillips Petroleum Company en 2002, **ConocoPhillips** se convirtió en la tercera compañía de energía integral más grande de Estados Unidos, con 160 000 millones de dólares en activos y 38 000 empleados. Como cualquier otra compañía en la industria, la administración de ConocoPhillips debe enfrentarse de manera continua con decisiones acerca de la asignación de capital de inversión limitado a los diferentes proyectos de exploración petrolera de alto riesgo. Dichas decisiones tienen un alto efecto en la rentabilidad de la compañía.

A principios de los años noventa, la entonces compañía Phillips Petroleum Company se convirtió en líder de la industria en la aplicación de una compleja metodología de investigación de operaciones con el fin de dar soporte a estas decisiones mediante el desarrollo de un *paquete de software de análisis de decisiones* llamado DISCOVERY. La interface de usuario le permite a un geólogo o ingeniero modelar la incertidumbre asociada con un proyecto y después el software interpreta las entradas y construye un árbol de decisiones que muestra todos los nodos de decisiones (lo cual incluye las oportunidades para obtener información sísmica adicional) así como los nodos que intervinieron en el evento. Una característica fundamental del software radica en el uso de una *función de utilidad exponencial* (la cual se presentará en la sección 15.6) para incorporar las actitudes de la gerencia acerca de riesgo financiero. Se utiliza un cuestionario intuitivo con el fin de medir las preferencias de riesgo de la corporación con el objeto de poder determinar un valor apropiado del parámetro de tolerancia al riesgo de esta función utilitaria.

La administración utiliza dicho software para 1) evaluar los *proyectos de exploración petrolera* con una política consistente en cuanto a toma de riesgos, 2) valorar los proyectos en términos de su preferencia en general, 3) identificar el nivel de participación adecuado de la firma en estos proyectos y 4) conservarse dentro del presupuesto acordado.

Fuente: M. R. Walls, G. T. Morahan y J. S. Dyer: "Decision Analysis of Exploration Opportunities in the Onshore US at Phillips Petroleum Company", en *Interfaces*, **25**(6): 39-56, noviembre-diciembre de 1995. (En el sitio en internet de este libro [www.mhhe.com/hillier] se puede encontrar una liga hacia este artículo.)

Criterio del pago máximo

Si el problema del tomador de decisiones se viera como *un juego contra la naturaleza*, la teoría de juegos diría que se seleccionara la opción de acuerdo con el *criterio minimax* (descrito en la sección 14.2). Desde el punto de vista del jugador 1 (tomador de decisiones), un nombre más adecuado para este criterio es el *del pago máximo*, como se resume a continuación.

> **Criterio del pago máximo:** Para cada opción posible, encuentre el *pago mínimo* sobre todos los estados posibles de la naturaleza. Después, encuentre el *máximo* de estos pagos mínimos. Elija la opción cuyo pago mínimo corresponde a este máximo.

En la tabla 15.3 se muestra la aplicación de este criterio al ejemplo prototipo. Entonces, como el pago mínimo por vender (90) es más grande que el de perforar (−100), se elegirá la segunda acción (vender el terreno).

El razonamiento que fundamenta este criterio es que proporciona la *mejor garantía* del pago que se obtendrá. Sin que importe cuál sea el estado de la naturaleza verdadero en el caso del ejemplo, el pago por vender el terreno no puede ser menor que 90, lo cual proporciona la mejor garantía disponible. De esta forma, el criterio toma el punto de vista pesimista de que, sin importar qué acción se elija, es posible que ocurra el peor estado de la naturaleza para esta opción, por lo que debe elegirse la opción que proporcione el mejor pago para el peor estado de la naturaleza.

Este razonamiento es bastante válido cuando se compite contra un oponente racional y malévolo. Sin embargo, este criterio casi no se usa en juegos contra la naturaleza puesto que es demasiado conservador en este contexto. En efecto, supone que la naturaleza es un oponente consciente

■ **TABLA 15.2 Tabla de pagos para la formulación de análisis de decisiones del problema de Goferbroke Co.**

	Estado de la naturaleza	
Alternativa	**Petróleo**	**Seco**
1. Perforar en busca de petróleo	700	−100
2. Vender terreno	90	90
Probabilidad *a priori*	0.25	0.75

TABLA 15.3 Aplicación del criterio de pagos maximin en el primer problema de la Goferbroke Co.

	Estado de la naturaleza			
Alternativa	**Petróleo**	**Seco**	**Mínimo**	
1. Perforar en busca de petróleo	700	−100	−100	
2. Vender terreno	90	90	90	← Valor máximo
Probabilidad *a priori*	0.25	0.75		

que quiere infligir al tomador de decisiones todo el daño que le sea posible. La naturaleza no es un oponente malévolo y el tomador de decisiones no necesita enfocar su atención sólo en el peor pago de cada acción. Esto es cierto en especial cuando el peor de los pagos se genera por un estado de la naturaleza relativamente poco probable.

Así, este criterio es de interés en general para un tomador de decisiones muy precavido.

Criterio de la máxima posibilidad

El siguiente criterio se refiere al estado *más probable* de la naturaleza, como se resume a continuación.

> **Criterio de la máxima posibilidad:** Identifique el estado más probable de la naturaleza (aquel que tiene la probabilidad *a priori* más grande). Para este estado de la naturaleza, encuentre la opción con el máximo pago. Elija esta alternativa de decisión.

Cuando se aplica este criterio al ejemplo, en la tabla 15.4 se indica que el estado *Seco* tiene la mayor probabilidad *a priori*. En la columna Seco, el pago máximo corresponde a la venta, por lo que la elección es vender el terreno.

Lo que más llama la atención de este criterio es que el estado más importante de la naturaleza es el que tiene más posibilidades de ocurrir, de manera que la acción elegida es la mejor para este estado. Basar la decisión en el supuesto de que este estado de la naturaleza ocurrirá tiende a dar mejor posibilidad de un resultado favorable que suponer cualquier otro. Más aún, el criterio no se basa en estimaciones subjetivas cuestionables de las probabilidades de los respectivos estados de la naturaleza que no sean el hecho de identificar el estado más probable.

La mayor desventaja del criterio es que hace caso omiso de mucha información relevante. No considera otro estado de la naturaleza distinto al más probable. En un problema con muchos estados posibles, la posibilidad de que ocurra el más probable puede ser bastante pequeña, por lo que centrarse sólo en él es bastante riesgoso. Aun en el ejemplo, donde la probabilidad *a priori* del estado *Seco* es 0.75, este criterio pasa por alto el atractivo pago de 700 que supone perforar y encontrar petróleo. De hecho, el criterio no permite jugar con un pago alto que tiene una probabilidad baja, no importa qué tan atractivo sea tomar el riesgo.

TABLA 15.4 Aplicación de la regla de máxima probabilidad en el primer problema de la Goferbroke Co.

	Estado de la naturaleza			
Alternativa	**Petróleo**	**Seco**		
1. Perforar en busca de petróleo	700	−100	−100	
2. Vender terreno	90	90	90	← Valor máximo en esta columna
Probabilidad *a priori*	0.25	0.75		
		↑ Máximo		

Regla de decisión de Bayes[1]

El tercer criterio que se usa con más frecuencia es la *regla de decisión de Bayes*, descrita como:

> **Regla de decisión de Bayes:** Se utilizan las mejores estimaciones disponibles de las probabilidades de los respectivos estados de la naturaleza —en este momento las probabilidades *a priori*—, para calcular el valor esperado del pago de cada opción posible. Se elige la opción con el máximo pago esperado.

En el ejemplo prototipo, estos pagos esperados se calculan directamente de la tabla 15.2 de la siguiente manera:

$$\begin{aligned} E[\text{Pago (perforar)}] &= 0.25(700) + 0.75(-100) \\ &= 100. \\ E[\text{Pago (venta)}] &= 0.25(90) + 0.75(90) \\ &= 90. \end{aligned}$$

Como 100 es mayor que 90, la opción que se debe seleccionar es perforar en busca de petróleo.

Observe que esta elección es contraria a elegir la venta del terreno que se obtuvo con los dos criterios anteriores.

La gran ventaja de la regla de decisión de Bayes es que incorpora toda la información disponible, incluso todos los pagos y las mejores estimaciones disponibles de las probabilidades de los respectivos estados de la naturaleza.

En ocasiones se argumenta que estas estimaciones de las probabilidades son necesariamente subjetivas y por lo tanto no se puede confiar en ellas. No existe una manera exacta, ni siquiera en términos de probabilidades, de predecir el futuro, lo cual incluye un estado de la naturaleza futuro. Este argumento tiene cierta validez. Deberá evaluarse qué tan razonables son estas estimaciones de las probabilidades en el caso de cada situación individual.

No obstante, en muchas circunstancias, la experiencia y la evidencia actual permiten desarrollar estimaciones razonables de las probabilidades. Usar esta información deberá proporcionar un fundamento más sólido para tomar una buena decisión que se ignora. Aún más, con frecuencia se puede llevar a cabo una experimentación para mejorar estas estimaciones, tema que se explicará en la siguiente sección. Por lo tanto, en este capítulo sólo se usará la regla de decisión de Bayes.

Para evaluar el efecto de posibles inexactitudes de las probabilidades *a priori*, suele ser útil realizar un análisis de sensibilidad, como se describe a continuación.

Análisis de sensibilidad con la regla de decisión de Bayes

Es común el uso del análisis de sensibilidad con varias aplicaciones de investigación de operaciones para estudiar el efecto si algunos números incluidos en el modelo matemático no son correctos. En este caso, el modelo matemático está representado por la tabla de pagos que se muestra en la tabla 15.2. Los números de esta tabla más cuestionables son las probabilidades *a priori*. Este estudio se centrará en el análisis de sensibilidad sobre estos números, aunque se puede aplicar un enfoque similar a los pagos dados en la tabla.

La suma de las dos probabilidades *a priori* debe ser 1, por lo que si se aumenta una de ellas, la otra debe disminuir de manera automática en la misma cantidad, y viceversa. La administración de Goferbroke cree que la posibilidad ideal de encontrar petróleo en el área debe estar entre 15 y 35%. En otras palabras, es posible que la probabilidad *a priori* verdadera de encontrar petróleo esté entre 0.15 y 0.35, de manera que encontrar que el terreno esté seco tendrá probabilidad *a priori* entre 0.85 y 0.65.

---

[1] El origen de este nombre es que con frecuencia se da crédito por él al Reverendo Thomas Bayes, un ministro inglés no conformista del siglo XVIII reconocido como filósofo y matemático. (La misma idea básica tiene raíces más antiguas en el campo de la economía.) Esta regla de decisión también suele llamarse criterio del *valor monetario esperado* (*VME*), no obstante que es un mal nombre para los casos en que la medida de pago es otra que el valor monetario (como en la sección 15.6).

Sea

p = probabilidad *a priori* de encontrar petróleo,

el pago esperado de perforar para cualquier valor de p es

$$\begin{aligned} E[\text{Pago (perforar)}] &= 700p - 100(1 - p) \\ &= 800p - 100. \end{aligned}$$

La línea inclinada en la figura 15.1 muestra la gráfica de este pago esperado contra p. Como el pago por la venta sería 90 para cualquier p, la línea horizontal de la figura 15.1 da E[Pago(venta)] contra p.

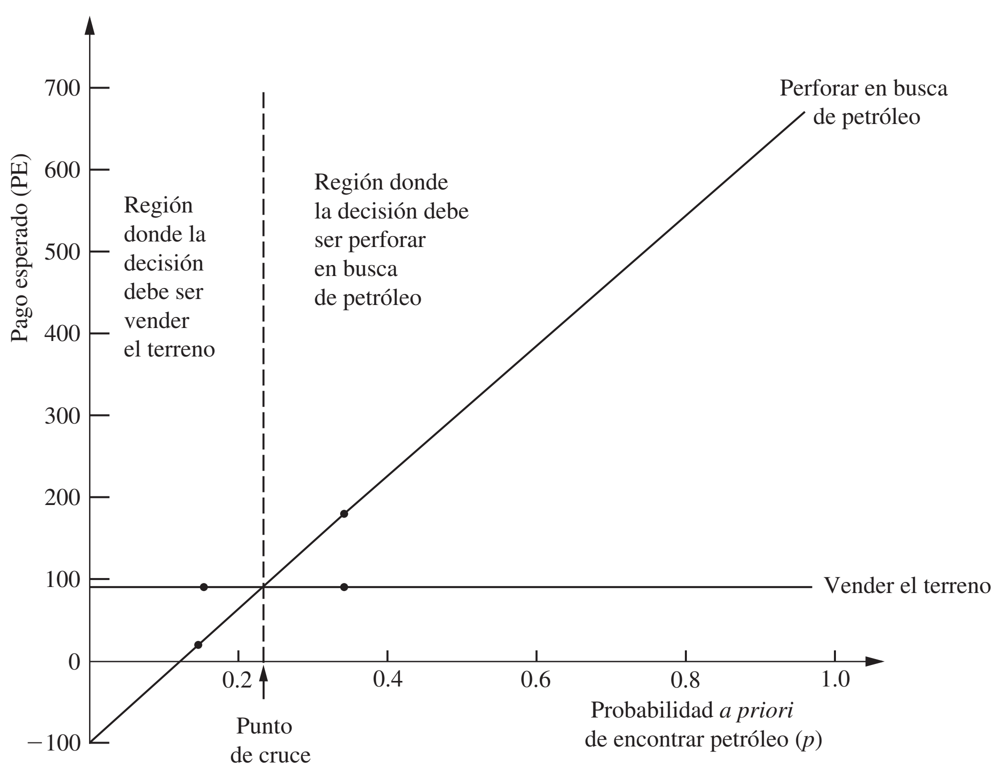
Los cuatro puntos de la figura 15.1 muestran el pago esperado para las dos alternativas de decisión cuando $p = 0.15$ o $p = 0.35$. Cuando $p = 0.15$, la decisión se inclina hacia la venta del terreno por un amplio margen (un pago esperado de 90 contra sólo 20 por la perforación). Sin embargo, cuando $p = 0.35$, la decisión es perforar con un margen amplio (pago esperado = 180 contra sólo 90 por la venta). Entonces, la decisión es muy *sensible* a la probabilidad *a priori* de encontrar petróleo. Este análisis de sensibilidad revela que es importante hacer algo más, si es posible, para desarrollar una estimación más precisa del valor verdadero de p.

El punto en la figura 15.1 donde se cruzan las dos rectas es el **punto de cruce** donde la decisión cambia de una alternativa (vender) a la otra (perforar) cuando la probabilidad *a priori* aumenta. Para encontrar este punto se establece

$$\begin{aligned} E[\text{Pago (perforar]} &= E[\text{Pago (venta)}] \\ 800p - 100 &= 90 \\ p &= \frac{190}{800} = 0.2375 \end{aligned}$$

Conclusión: Se debe vender el terreno si $p < 0.2375$.
Se debe perforar en busca de petróleo si $p > 0.2375$.

FIGURA 15.1
Gráfica del cambio del pago esperado para cada acción alternativa cuando cambia la probabilidad *a priori* de encontrar petróleo para el primer problema de Goferbroke Co.



Así, cuando se trate de refinar la estimación del valor verdadero de p, la pregunta clave es si éste es mayor o menor que 0.2375.

En otros problemas que tienen más de dos opciones alternativas, se puede aplicar el mismo tipo de análisis. La diferencia principal es que habrá más de dos rectas (una por alternativa) en la gráfica que corresponde a la figura 15.1. Sin embargo, la línea superior de cualquier valor de la probabilidad *a priori* aún indicará cuál alternativa debe elegirse. Con más de dos líneas puede haber más de un punto de cruce donde la decisión cambia de una alternativa a otra.

En la sección Worked Examples del sitio en internet de este libro se puede ver **otro ejemplo** en el que se realiza este tipo de análisis con tres alternativas de decisión. (Con este mismo ejemplo también se ilustra la aplicación de los tres criterios de decisión que se consideran en esta sección.)

En el caso de un problema con más de dos estados de la naturaleza posibles, el enfoque más directo es centrar el análisis de sensibilidad en sólo dos estados a la vez como ya se describió. De nuevo, esto implica investigar qué pasa cuando la probabilidad *a priori* de un estado aumenta mientras la del otro disminuye en la misma cantidad, y se mantienen fijas las probabilidades *a priori* de los estados restantes. Este procedimiento se repite para los pares de estados que se desee.

Quienes aplican teoría de decisiones suelen usar paquetes de software como ayuda en el análisis de sensibilidad, incluso la generación de gráficas. Por ejemplo, un complemento de Excel en el OR Courseware llamado *SensIt* está diseñado para realizar el análisis de sensibilidad con modelos probabilísticos como la regla de decisión de Bayes. En el sitio en internet de este libro se incluye la documentación completa de SensIt. En la sección 15.5 se describirá e ilustrará la aplicación de este complemento de Excel.

Debido a que la decisión que Goferbroke Co. debe tomar depende de manera tan crítica de la probabilidad verdadera de encontrar petróleo, debe considerarse de manera seria la realización de una investigación sismológica para estimar mejor esta probabilidad. Esta opción se explorará en las dos secciones siguientes.

15.3 TOMA DE DECISIONES CON EXPERIMENTACIÓN

Es frecuente hacer pruebas adicionales (experimentación) para mejorar las estimaciones preliminares de las probabilidades de los respectivos estados de la naturaleza dadas por las probabilidades *a priori.* Estas estimaciones mejoradas se llaman **probabilidades *a posteriori***.

Primero se actualizará el ejemplo de la Goferbroke Co. para incorporar la experimentación, después se describirá cómo obtener las probabilidades *a posteriori* y, por último, se analizará cómo decidir si vale la pena realizar la experimentación.

Continuación del ejemplo prototipo

Como se mencionó al final de la sección 15.1, una opción disponible antes de tomar una decisión es llevar a cabo una exploración sismológica del terreno para obtener una mejor estimación de la probabilidad de que haya petróleo. El costo es de 30 000 dólares.

Una exploración sismológica obtiene sondeos sísmicos que indican si la estructura geológica es favorable para la presencia de petróleo. Los resultados posibles de la exploración se dividen en las siguientes categorías:

SSD: sondeos sísmicos desfavorables; es poco probable encontrar petróleo.
SSF: sondeos sísmicos favorables; es bastante probable encontrar petróleo.

Con base en la experiencia, si hay petróleo, la probabilidad de sondeos sísmicos desfavorables es:

$$P(\text{SSD} \mid \text{Estado} = \text{Petróleo}) = 0.4, \qquad \text{y} \qquad P(\text{SSF} \mid \text{Estado} = \text{Petróleo}) = 1 - 0.4 = 0.6.$$

De igual manera, si no hay petróleo —es decir, si el verdadero estado de la naturaleza es *Seco*—, entonces la probabilidad de sondeos sísmicos desfavorables se estima en:

$$\text{K}(\text{SSD} \mid \text{Estado} = \text{Seco}) = 0.8, \qquad \text{y} \qquad \text{P}(\text{SSF} \mid \text{Estado} = \text{Seco}) = 1 - 0.8 = 0.2.$$

Más adelante se usarán estos datos para encontrar las probabilidades *a posteriori* de los respectivos estados de la naturaleza *dados* los sondeos sísmicos.

Probabilidades *a posteriori*

Ahora en términos generales, sea

n = número posible de estados de la naturaleza;

$P(\text{Estado} = \text{estado } i)$ = probabilidad *a priori* de que el estado de la naturaleza verdadero sea el estado i, para $i = 1, 2, \ldots, n$;

Resultado = resultado de la experimentación (una variable aleatoria);

Resultado j = un valor posible del resultado;

$P(\text{Estado} = \text{estado } i \mid \text{Resultado} = \text{resultado } j)$ = probabilidad *a posteriori* de que el estado de la naturaleza verdadero sea el estado i, dado que Resultado = resultado j, para $i = 1, 2, \ldots, n$.

La pregunta que se hace en este momento es la siguiente:

> Dados $P(\text{Estado} = \text{estado } i)$ y $P(\text{Resultado} = \text{resultado } j \mid \text{Estado} = \text{estado } i)$, para $i = 1, 2, \ldots, n$, ¿cuál es el valor de $P(\text{Estado} = \text{estado } i \mid \text{Resultado} = \text{resultado } j)$?

La respuesta se obtiene con la combinación de las siguientes fórmulas estándar de teoría de probabilidad:

$$P(\text{Estado} = \text{estado } i \mid \text{Resultado} = \text{resultado } j) = \frac{P(\text{Estado} = \text{estado } i, \text{Resultado} = \text{resultado } j)}{P(\text{Resultado} = \text{resultado } j)}$$

$$P(\text{Resultado} = \text{resultado } j) = \sum_{k=1}^{n} P(\text{Estado} = \text{estado } k, \text{Resultado} = \text{resultado } j)$$

$$P(\text{Estado} = \text{estado } i, \text{Resultado} = \text{resultado } j) = P(\text{Resultado} = \text{resultado } j \mid \text{Estado} = \text{estado } i)\, P(\text{Estado} = \text{estado } i).$$

Por lo tanto, para cada $i = 1, 2, \ldots, n$, la fórmula deseada para la probabilidad *a posteriori* correspondiente es

$$P(\text{Estado} = \text{estado } i \mid \text{Resultado} = \text{resultado } j) = \frac{P(\text{Resultado} = \text{resultado } j \mid \text{Estado} = \text{estado } i)\, P(\text{Estado} = \text{estado } i)}{\sum_{k=1}^{n} P(\text{Resultado} = \text{resultado } j \mid \text{Estado} = \text{estado } k)\, P(\text{Estado} = \text{estado } k)}$$

(Esta fórmula con frecuencia se conoce como **teorema de Bayes** porque fue desarrollada por Thomas Bayes, el mismo matemático del siglo XVIII a quien se da crédito por desarrollar la regla de decisión que lleva su nombre.)

Ahora, se aplicará esta fórmula al ejemplo prototipo. Si el resultado de la exploración son sondeos sísmicos desfavorables (SSD), entonces las probabilidades *a posteriori* son

$$P(\text{Estado} = \text{Petróleo} \mid \text{Resultado} = \text{SSD}) = \frac{0.4(0.25)}{0.4(0.25) + 0.8(0.75)} = \frac{1}{7},$$

$$P(\text{Estado} = \text{Seco} \mid \text{Resultado} = \text{SSD}) = 1 - \frac{1}{7} = \frac{6}{7}.$$

De manera similar, si en la exploración se obtienen sondeos sísmicos favorables (SSF), entonces

$$P(\text{Estado} = \text{Petróleo} \mid \text{Resultado} = \text{SSF}) = \frac{0.6(0.25)}{0.6(0.25) + 0.2(0.75)} = \frac{1}{2},$$

$$P(\text{Estado} = \text{Seco} \mid \text{Resultado} = \text{SSF}) = 1 - \frac{1}{2} = \frac{1}{2}.$$

Recuadro de aplicación

El **Workers'Compensation Board (WCB) de la Columbia Británica, Canadá**, es responsable de los intereses de los trabajadores y empleados de esta provincia en cuanto a salud y seguridad, rehabilitación y compensación. Por ello, brinda servicio a más de 165 000 empleadores que dan trabajo a alrededor de 1.8 millones de trabajadores en Columbia Británica y gasta alrededor de 1 000 millones de dólares estadounidenses en rehabilitación y compensaciones.

Un factor clave en el control de los costos del Consejo consiste en identificar las reclamaciones de incapacitados en el corto plazo que representen un potencialmente elevado riesgo financiero de convertirse en una reclamación a largo plazo mucho más costosa a menos que haya una *intervención del manejo de reclamaciones intensiva* y *temprana* para brindar el tratamiento y la rehabilitación médica necesaria. La pregunta fue cómo identificar con precisión dichas reclamaciones de alto riesgo de tal manera que se minimicen los costos totales esperados por concepto de compensación de reclamaciones e intervención del manejo de reclamaciones.

Se formó un equipo de investigación de operaciones para estudiar este problema mediante la *aplicación del análisis de decisiones*. Por cada una de las numerosas categorías de reclamaciones por lesiones y con base en la naturaleza de la misma, el sexo y la edad del trabajador, etc. se utilizó un *árbol de decisiones* para evaluar si dicha categoría debe clasificarse como de bajo riesgo (es decir, que no requiere intervención) o de alto riesgo (que requiere intervención), en función del grado de severidad de la lesión. Por cada categoría, se realizó un cálculo del punto de corte del número crítico de días por reclamaciones de incapacitados en el corto plazo que dispararía la intervención del manejo de reclamaciones, de tal manera que minimice el costo esperado de los pagos e intervención de la reclamación. Un aspecto clave en la realización de este cálculo fue la evaluación de la probabilidad posterior de que la reclamación se convirtiera en una reclamación por incapacidad en el largo plazo, dado el número de días pagados por reclamaciones de incapacidad en el corto plazo.

Esta aplicación del análisis de decisiones mediante árboles de decisión *ahorra al WCB alrededor* de **4 millones de dólares** estadounidenses *al año* a la vez que permite que algunos trabajadores lesionados se puedan reincorporar más pronto a su trabajo.

Fuente: E. Urbanovich, E. E. Young, M. L. Puterman y S. O. Fattedad: "Early Detection of High-Risk Claims at the Workers' Compensation Board of British Columbia", en *Interfaces*, **33**(4):15-26, julio-agosto de 2003. (En el sitio en internet de este libro —www.mhhe.com/hillier— se proporciona una liga hacia este artículo.)

En el **diagrama de árbol de probabilidad** de la figura 15.2 se muestra una manera conveniente de organizar estos cálculos de una manera intuitiva. Las probabilidades *a priori* en la primera columna y las probabilidades condicionales en la segunda son parte de los datos de entrada del problema. Cuando se multiplica cada probabilidad de la primera columna por una probabilidad de la segunda se obtiene la probabilidad conjunta correspondiente de la tercera columna. Cada probabilidad conjunta se convierte en el numerador del cálculo correspondiente de la probabilidad *a posteriori* en la cuarta columna. Al acumular las probabilidades conjuntas con los mismos resultados —como se muestra abajo del árbol en la figura— se obtiene el denominador de cada probabilidad *a posteriori* con este resultado. (Si el lector desea ver **otro ejemplo** de la utilización de un diagrama de árbol de probabilidades para determinar las probabilidades *a posteriori*, uno de éstos se incluye en la sección Worked Examples del sitio en internet de este libro.)

El OR Courseware incluye también una plantilla de Excel para calcular estas probabilidades *a posteriori*, como la que se muestra en la figura 15.3.

Una vez completos estos cálculos, se puede aplicar la regla de decisión de Bayes igual que antes, donde las probabilidades *a posteriori* sustituyen a las probabilidades *a priori*. De nuevo, si se usan los pagos (en miles de dólares) dados en la tabla 15.2 y se resta el costo de la experimentación, se obtienen los resultados que se muestran a continuación.

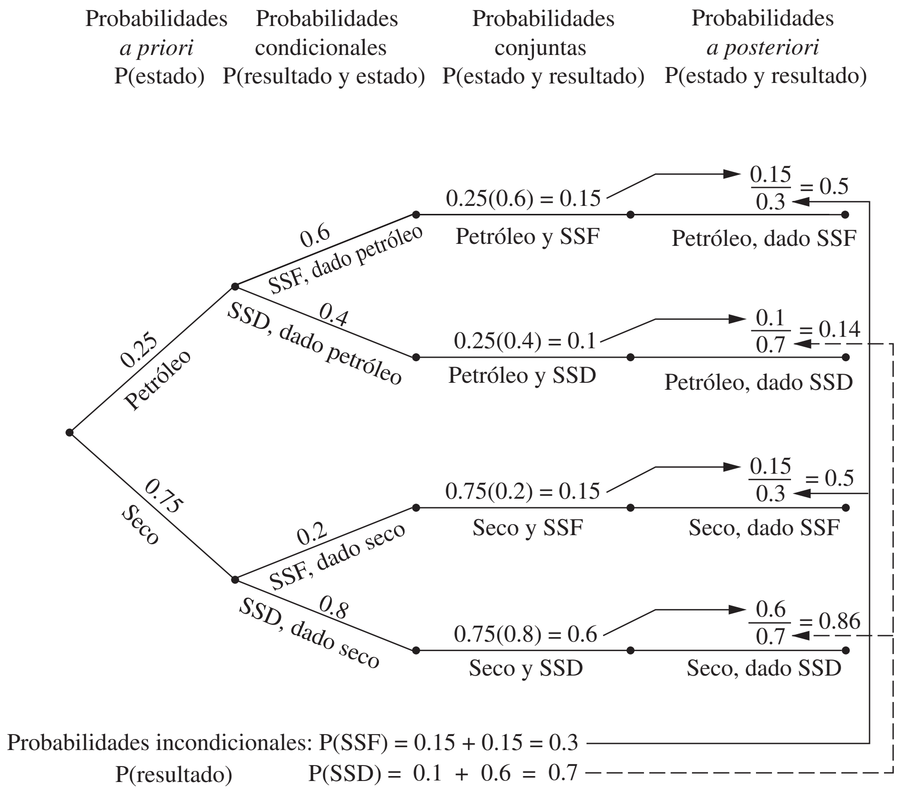
Pago esperado si el resultado es un sondeo desfavorable (SSD):

$$E[\text{Pago (perforar} \mid \text{Resultado} = \text{SSD)}] = \frac{1}{7}(700) + \frac{6}{7}(-100) - 30$$

$$= -15.7.$$

$$E[\text{Pago (venta} \mid \text{Resultado} = \text{SSD)}] = \frac{1}{7}(90) + \frac{6}{7}(90) - 30$$

$$= 60.$$



■ **FIGURA 15.2**
Diagrama de árbol de probabilidades del problema completo de Goferbroke Co. que muestra todas las probabilidades que conducen al cálculo de cada probabilidad *a posteriori* del estado de la naturaleza dado el resultado del sondeo sísmico.

■ **FIGURA 15.3**
Plantilla de *probabilidades a posteriori* del OR Courseware que permite el cálculo eficiente de probabilidades, como se ilustra para el problema completo de la Goferbroke Co.

	A	B	C	D	E	F	G	H
1	**Template for Posterior Probabilities**							
2								
3		**Data:**		P(Finding I State)				
4		State of	Prior	Finding				
5		Nature	Probability	FSS	USS			
6		Oil	0.25	0.6	0.4			
7		Dry	0.75	0.2	0.8			
8								
9								
10								
11								
12		**Posterior**		P(State I Finding)				
13		**Probabilities:**		State of Nature				
14		Finding	P(Finding)	Oil	Dry			
15		FSS	0.3	0.5	0.5			
16		USS	0.7	0.14286	0.85714			
17								
18								
19								

	B	C	D
12	**Posterior**		P(State I Finding)
13	**Probabilities:**		State of Nature
14	Finding	P(Finding)	=B6
15	=D5	=SUMPRODUCT(C6:C10,D6:D10)	=C6*D6/SUMPRODUCT(C6:C10,D6:D10)
16	=E5	=SUMPRODUCT(C6:C10,E6:E10)	=C6*E6/SUMPRODUCT(C6:C10,E6:E10)
17	=F5	=SUMPRODUCT(C6:C10,F6:F10)	=C6*F6/SUMPRODUCT(C6:C10,F6:F10)
18	=G5	=SUMPRODUCT(C6:C10,G6:G10)	=C6*G6/SUMPRODUCT(C6:C10,G6:G10)
19	=H5	=SUMPRODUCT(C6:C10,H6:H10)	=C6*H6/SUMPRODUCT(C6:C10,H6:H10)

TABLA 15.5 Política óptima con experimentación, de acuerdo con la regla de Bayes, para el problema de la Goferbroke Co.

Resultado del sondeo	Acción óptima	Pago esperado excluyendo el costo de la exploración	Pago esperado incluyendo el costo de la exploración
SSD	Vender el terreno	90	60
SSF	Perforar en busca de petróleo	300	270

Pago esperado si el resultado es un sondeo favorable (SSF):

$$E[\text{Pago (perforar} \mid \text{Resultado} = \text{SSF)}] = \frac{1}{2}(700) + \frac{1}{2}(-100) - 30$$

$$= 270.$$

$$E[\text{Pago (venta} \mid \text{Resultado} = \text{SSF)}] = \frac{1}{2}(90) + \frac{1}{2}(90) - 30$$

$$= 60.$$

Como el objetivo es maximizar el pago esperado, estos resultados conducen a la política óptima que se muestra en la tabla 15.5.

Sin embargo, lo que este análisis no resuelve es si vale la pena gastar 30 000 dólares para llevar a cabo la experimentación (la exploración de sismología). Tal vez sería mejor evitar este importante gasto y sólo usar la solución óptima sin experimentación, es decir, perforar en busca de petróleo con un pago esperado de 100 000 dólares. Este aspecto se estudiará en seguida.

El valor de la experimentación

Antes de realizar cualquier experimento debe determinarse su valor potencial. Se presentan aquí dos métodos complementarios para evaluar su valor potencial.

Cuando se aplica el primer método se supone (de manera poco realista) que la experimentación eliminará *toda* la incertidumbre sobre cuál es el estado verdadero de la naturaleza y después se hace un cálculo rápido sobre cuál sería la *mejora del pago esperado* (ignora el costo de experimentación). Esta cantidad, llamada *valor esperado de la información perfecta*, proporciona una *cota superior* del valor potencial del experimento. Por lo tanto, si esta cota superior es menor que el costo del experimento, en definitiva éste no debe llevarse a cabo.

Sin embargo, si esta cota superior excede el costo de la experimentación, debe usarse el segundo método (más lento). Este método calcula la mejora *real* del pago esperado (pasa por alto el costo de experimentación) que resultaría de realizar el experimento. La comparación de esta mejora (llamada *valor esperado de la experimentación*) con el costo indica si el experimento debe llevarse a cabo.

Valor esperado de la información perfecta. Suponga ahora que el experimento puede identificar de manera definitiva cuál es el verdadero estado de la naturaleza y proporcionar con esto información "perfecta". Cualquiera que sea el estado de la naturaleza identificado, se debe elegir la opción con el máximo pago para ese estado. No se sabe de antemano cuál estado se identificará, por lo que el cálculo del pago esperado con la información perfecta (sin el costo de experimentación) requiere ponderar el pago máximo para cada estado de la naturaleza con la probabilidad *a priori* de ese estado.

Al final de la tabla 15.6 se presenta este cálculo para el problema completo de Goferbroke Co., donde el valor esperado de la información perfecta es 242.5. Entonces, si Goferbroke puede saber, antes de elegir su opción, si hay petróleo en el terreno, el pago esperado por ahora (antes de adquirir esta información) sería de 242 500 dólares (sin el costo del experimento para generar la información).

■ **TABLA 15.6 Pago esperado con información perfecta para el problema completo de la Goferbroke Co.**

	Estado de la naturaleza	
Alternativa	**Petróleo**	**Seco**
1. Perforar en busca del petróleo	700	−100
2. Perforar el terreno	90	90
Pago máximo	700	90
Probabilidad *a priori*	0.25	0.75
Pago esperado con información perfecta = 0.25(700) + 0.75(90) = 242.5		

Para evaluar si debe realizarse el experimento, se usa esta cantidad para calcular el valor esperado de la información perfecta.

El **valor esperado de la información perfecta**, abreviado **VEIP**, se calcula como

VEIP = pago esperado con información perfecta − pago esperado sin experimentación.[2]

Así como la experimentación casi nunca puede proporcionar información perfecta, el VEIP resulta ser una cota superior sobre el valor esperado de la experimentación.

Para este mismo ejemplo, en la sección 15.2 se encontró que el pago esperado sin experimentación (según la regla de decisión de Bayes) es 100. Por lo tanto,

$$\text{VEIP} = 242.5 - 100 = 142.5.$$

Como 142.5 excede por mucho a 30, el costo de la experimentación (el sondeo sísmico), puede valer la pena proceder con la exploración. Para confirmar esta decisión se estudiará un segundo método de evaluación del beneficio potencial de la experimentación.

Valor esperado de la experimentación. En lugar de sólo obtener la cota superior *del incremento de pago esperado* (sin el costo de la experimentación) debido a que se lleva a cabo la experimentación, ahora se realizará un poco más de trabajo para calcular, de manera directa, este incremento esperado. Esta cantidad se llama el *valor esperado de la experimentación*.

El cálculo de esta cantidad requiere primero calcular el pago esperado con experimentación (sin el costo del experimento). La obtención de esta última cantidad requiere hacer todo el trabajo descrito para encontrar todas las probabilidades *a posteriori*, la política óptima con experimentación y el pago esperado correspondiente —que excluye el costo del experimento— de cada resultado posible del experimento. Después, cada pago esperado debe ponderarse con la probabilidad del resultado correspondiente, es decir,

$$\text{Pago esperado con experimentación} = \sum_j P(\text{Resultado} = \text{resultado } j) \, E[\text{pago} \mid \text{Resultado} = \text{resultado } j],$$

donde la suma se toma sobre todos los valores posibles de j.

Para el ejemplo prototipo, ya se hizo todo el trabajo de obtener los términos del lado derecho de esta ecuación. Los valores de $P(\text{Resultado} = \text{resultado } j)$ para los dos resultados posibles de estudio —desfavorable (SSD) y favorable (SSF)— se calcularon abajo del árbol de probabilidad de la figura 15.2 como

$$P(\text{SSD}) = 0.7, \qquad P(\text{SSF}) = 0.3.$$

[2] El *valor de la información perfecta* es una variable aleatoria igual al costo con la información perfecta menos el costo sin experimentación. VEIP es el valor esperado de esta variable aleatoria.

Para obtener la política óptima con experimentación, el pago esperado (sin el costo del estudio de sismología) de cada resultado se obtuvo en la tercera columna de la tabla 15.5 de la siguiente manera

$$E(\text{Pago} \mid \text{Resultado} = \text{SSD}) = 90,$$
$$E(\text{Pago} \mid \text{Resultado} = \text{SSF}) = 300.$$

Con estos números,

$$\begin{aligned}\text{Pago esperado con experimentación} &= 0.7(90) + 0.3(300)\\ &= 153.\end{aligned}$$

En este momento se puede calcular el valor esperado de la experimentación.

El **valor esperado de la experimentación**, abreviado VEE, se calcula como:

VEE = pago esperado con experimentación – pago esperado sin experimentación.

Entonces, VEE identifica el valor potencial de la experimentación.

En el caso de la Goferbroke Co.

$$\text{VEE} = 153 - 100 = 53.$$

Como este valor excede a 30, que es el costo de llevar a cabo un sondeo sísmico detallado (en miles de dólares), la experimentación debe realizarse.

15.4 ÁRBOLES DE DECISIÓN

Un árbol de decisión proporciona una forma para *desplegar visualmente* el problema y después *organizar el trabajo de cálculos* que se describió en las secciones anteriores. Estos árboles de decisión son en especial útiles cuando debe tomarse una *serie de decisiones*.

Construcción del árbol de decisión

El ejemplo prototipo incluye una serie de dos decisiones:

1. ¿Debe llevarse a cabo un sondeo sísmico antes de elegir una acción?
2. ¿Qué acción debe elegirse (perforar o vender el terreno)?

El árbol de decisión correspondiente (antes de llevar a cabo los cálculos) se muestra en la figura 15.4.

Los puntos de ramificación del árbol de decisión se conocen como **nodos** y los arcos se llaman **ramas**.

Un **nodo de decisión**, representado por un cuadrado, indica que en ese punto del proceso debe tomarse una decisión. Un **nodo de probabilidad**, representado por un círculo, indica que en ese punto ocurre un evento aleatorio.

Así, en la figura 15.4, la primera decisión se representa por el nodo de decisión *a*. El nodo *b* es un nodo de probabilidad que representa el evento aleatorio del resultado del sondeo sísmico. Las dos ramas que salen de *b* representan los dos resultados posibles de la exploración. Después es necesario tomar la segunda decisión (nodos *c*, *d* y *e*) con sus opciones posibles. Si la decisión es perforar, entonces se llega a otro nodo de probabilidad (los nodos *f*, *g* y *h*), donde las dos ramas corresponden a los dos estados posibles de la naturaleza.

Observe que la trayectoria seguida para llegar a cualquier pago (excepto la inferior) está determinada tanto por la decisión tomada como por los eventos aleatorios que están fuera del control del tomador de decisiones. Ésta es una característica de los problemas que se estudia en el análisis de decisiones.

El siguiente paso en la construcción del árbol de decisión es insertar números en el árbol como se muestra en la figura 15.5. Los números abajo y arriba de las ramas que *no* están entre paréntesis son flujos de efectivo (en miles de dólares) que ocurren en esas ramas. En el caso de cada trayectoria

Recuadro de aplicación

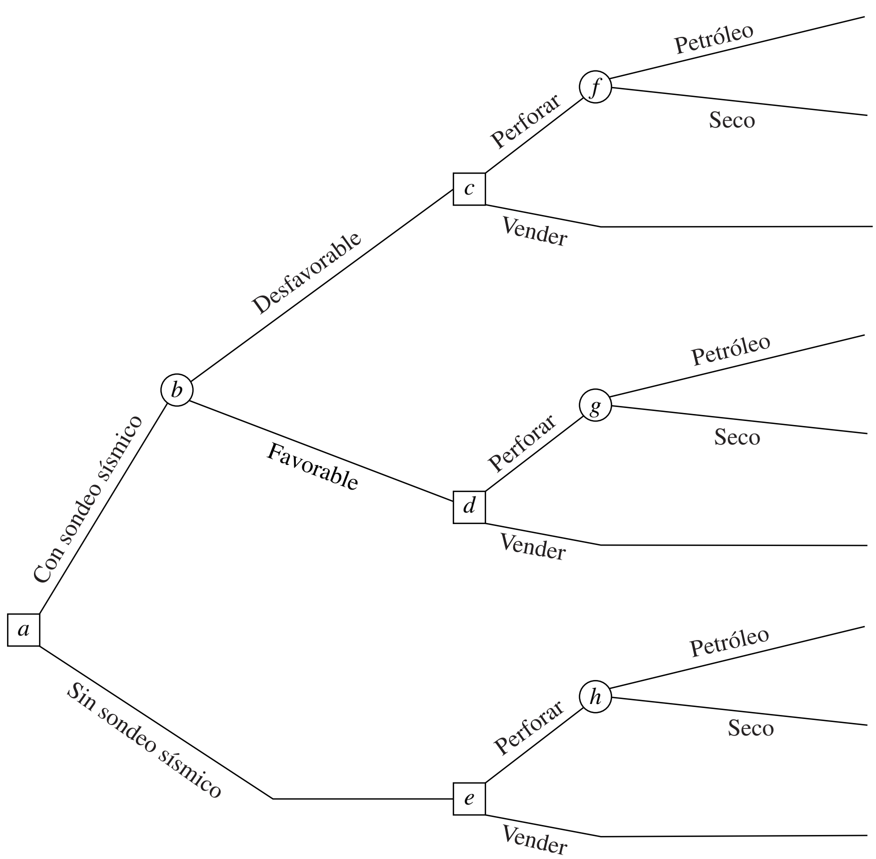
Westinghouse Science and Technology Center constituye la división de investigación y desarrollo (R&D) más importante de Westinghouse Electric Company para el desarrollo de nueva tecnología. El proceso de evaluación de proyectos de R&D para decidir cuáles deberán ponerse en marcha y después cuáles deben continuar a medida que se presentan avances (o no se presentan) es una tarea muy demandante para la administración debido a la gran incertidumbre y a los extensos periodos que implican. La fecha de lanzamiento real de una tecnología embriónica puede durar años o inclusive décadas desde su planteamiento como una propuesta discreta de R&D para investigar el potencial de dicha tecnología.

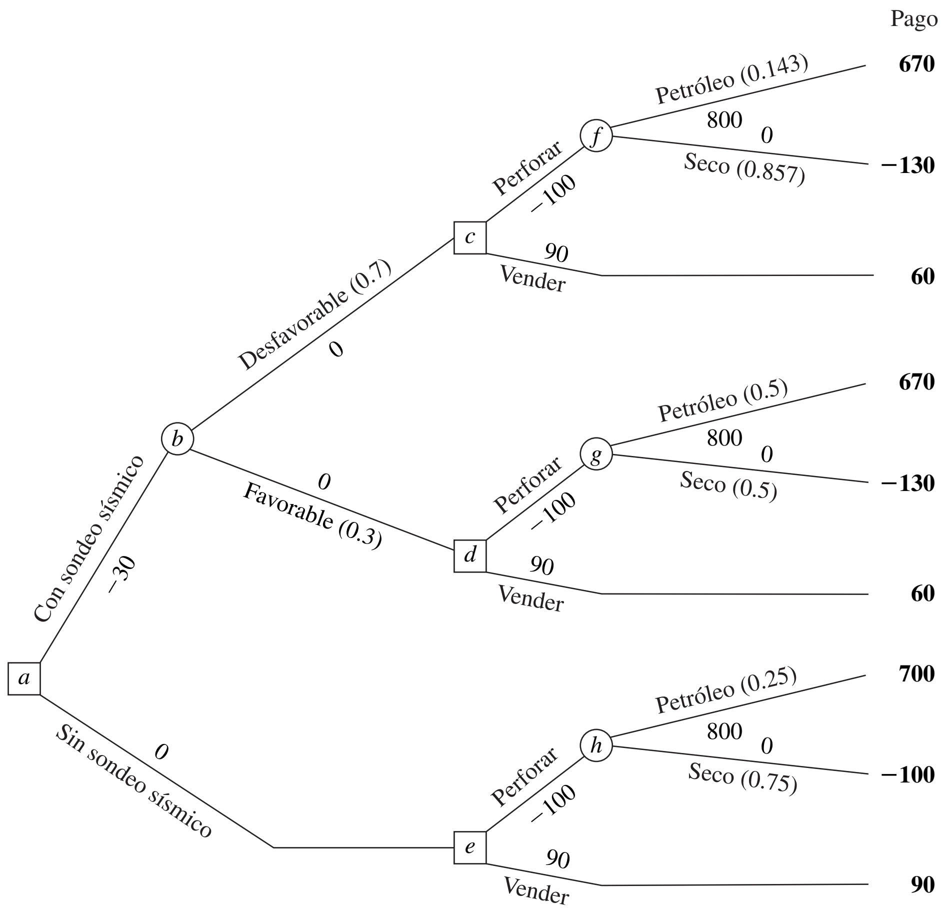
A medida que el Centro experimentó una presión cada vez mayor por reducir costos y crear tecnología de alto impacto en el corto plazo, su administrador financió un proyecto de investigación de operaciones con el fin de mejorar este proceso de evaluación. El grupo de investigación de operaciones desarrolló un método en forma de *árbol de decisiones* para analizar toda propuesta de R&D a la vez que consideraba la secuencia completa de puntos de decisión claves. El primer punto de decisión consistió en la conveniencia de financiar el proyecto embriónico propuesto para el primer año. En el caso de que se alcanzaran los objetivos técnicos iniciales, el segundo punto de decisión sería continuar financiando el proyecto por algún tiempo. Este proceso se podría repetir una y otra vez. Si se lograban los objetivos técnicos finales, el punto de decisión siguiente sería el prelanzamiento del producto debido a que la innovación aún cumple con los objetivos estratégicos del negocio. Si se cumple un objetivo estratégico, el punto de decisión final sería determinar la conveniencia de comercializar la innovación en ese momento o retrasarla o abandonar el proyecto. Un *árbol de decisiones* con una serie de nodos de decisión y nodos de eventos proporcionará una manera natural de describir y analizar dicho proyecto de R&D.

Fuente: R. K. Perdue, W. J. McAllister, P. V. King y B. G. Berkey: "Valuation of R and D Projects Using Options Pricing and Decision Analysis Models" en *Interfaces*, **29**(6): 57-74, noviembre-diciembre de 1999. (En el sitio en internet de este libro —www.mhhe.com/hillier— se proporciona una liga hacia este artículo.)

a través del árbol desde el nodo *a* hasta las ramas terminales, estos números se suman para obtener el pago total que resulta, que se muestra en negritas a la derecha de cada rama. El último conjunto de números es el de las probabilidades de los eventos. En particular, como cada rama que sale de un nodo de probabilidad representa un evento aleatorio posible, la probabilidad de este evento que ocurre a partir de él se encuentra entre paréntesis junto con esta rama. En el nodo de probabilidad

FIGURA 15.4
Árbol de decisión para el problema completo de la Goferbroke Co. (antes de incluir números).





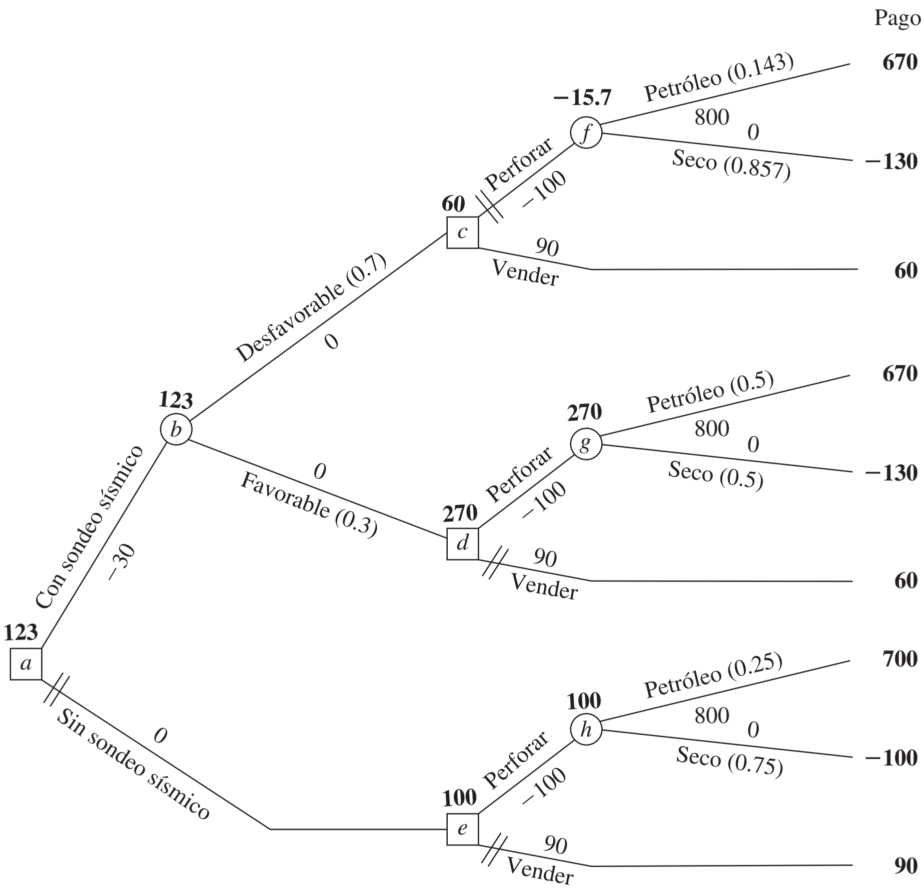
■ **FIGURA 15.5**
Árbol de decisión de la figura 15.4 después de agregar las probabilidades de los eventos aleatorios y los pagos.

h, las probabilidades son las *probabilidades a priori* de los estados de la naturaleza, puesto que en este caso no se ha realizado un sondeo sísmico para obtener más información. Sin embargo, los nodos *f* y *g* salen de una decisión de explorar (y después perforar). Entonces, las probabilidades de estos nodos son las *probabilidades a posteriori* de los estados de la naturaleza, dados los resultados del estudio, y los números se dan en las figuras 15.2 y 15.3. Por último, se tienen dos ramas que salen del nodo de probabilidad *b*. Los números son las probabilidades de los resultados del estudio, Favorable (SSF) o Desfavorable (SSD), como se presentan abajo del diagrama del árbol de probabilidad en la figura 15.2 o en las celdas C15:C16 de la figura 15.3.

Realización del análisis

Una vez construido el árbol de decisión, con sus números, se puede analizar el problema con el siguiente procedimiento:

1. Inicie en el lado derecho del árbol de decisión y muévase a la izquierda una columna a la vez. En cada columna, realice el paso 2 o el 3 según si los nodos en esa columna son de probabilidad o de decisión.
2. Para cada nodo de probabilidad, calcule su *pago esperado*. Para ello, debe multiplicar el pago esperado de cada rama (dado en **negritas** a la derecha) por la probabilidad de esa rama y después sumar ambos productos. Registre esta cantidad esperada en cada nodo de decisión en **negritas** junto al nodo y designe esta cantidad como el pago esperado de la rama que conduce a este nodo.
3. En cada nodo, compare los pagos esperados de sus ramas y seleccione la alternativa cuya rama tenga el mayor pago esperado. En cada caso, registre la elección en el árbol de decisión con una doble raya como barrera en las ramas rechazadas.



■ **FIGURA 15.6**
Árbol de decisión final que registra el análisis del problema completo de la Goferbroke Co. cuando se realizan pagos en dinero.

Para comenzar el procedimiento, considere la columna de nodos de la derecha, es decir, los nodos de probabilidad *f*, *g* y *h*. Si aplica el paso 2, sus pagos esperados (PE) serán

$$\text{PE} = \frac{1}{7}(670) + \frac{6}{7}(-130) = -15.7, \qquad \text{para el nodo } f,$$

$$\text{PE} = \frac{1}{2}(670) + \frac{1}{2}(-130) = 270, \qquad \text{para el nodo } g,$$

$$\text{PE} = \frac{1}{4}(700) + \frac{3}{4}(-100) = 100, \qquad \text{para el nodo } h.$$

Estos pagos esperados se colocan arriba de estos nodos, como se muestra en la figura 15.6.

A continuación se hace un movimiento una columna a la izquierda, que consiste en los nodos *c*, *d* y *e*. El pago esperado de una rama que conduce a un nodo de probabilidad se registra en **negritas** arriba de ese nodo. Por lo tanto, el paso 3 se puede aplicar de la siguiente manera:

Nodo *c*: la alternativa de perforar tiene PE $= -15.7$,
la alternativa de vender tiene PE $= 60$.
$60 > -15.7$, de manera que se elige la alternativa de vender.
Nodo *d*: la alternativa de perforar tiene PE $= 270$,
la alternativa de vender tiene PE $= 60$.
$270 > 60$, de manera que se elige la alternativa de perforar.
Nodo *e*: la alternativa de perforar tiene PE $= 100$,
la alternativa de vender tiene PE $= 90$.
$100 > 90$, de manera que se elige la alternativa de perforar.

El pago esperado de cada alternativa elegida se registra en **negritas** arriba del nodo de decisión, como se mostró en la figura 15.6. La alternativa seleccionada también se indica con la inserción de una doble raya como una barrera en cada rama rechazada.

Después, el proceso se mueve una columna más hacia la izquierda hasta el nodo *b*. Como es un nodo de probabilidad, debe aplicarse el paso 2 del procedimiento. El pago esperado de cada rama se registra arriba del nodo de decisión que sigue. Por lo tanto, el pago esperado es

$$\text{PE} = 0.7(60) + 0.3(270) = 123, \qquad \text{del nodo } b,$$

como se registra en dicho nodo en la figura 15.6.

Por último, el proceso se mueve al nodo de la izquierda *a*, un nodo de decisión. Con la aplicación del paso 3 se obtiene

Nodo *a*: Realizar el sondeo sísmico tiene PE = 123.
No realizar el sondeo tiene PE = 100.
$123 > 100$, de manera que se elige realizar el sondeo sísmico.

Este pago esperado de 123 se registra arriba del nodo y con la doble raya que indica la rama rechazada, como se muestra en la figura 15.6.

Este procedimiento se ha movido de derecha a izquierda para el análisis. Sin embargo, al completar el árbol, el tomador de decisiones ahora puede leerlo de izquierda a derecha para ver el avance real de los eventos. Las rayas dobles han cerrado las trayectorias no deseadas. Entonces, dados los pagos de los eventos finales en el lado derecho, la *regla de decisión de Bayes* dice que se siga sólo las trayectorias abiertas de izquierda a derecha para lograr el pago esperado mayor posible.

Al seguir las trayectorias abiertas de izquierda a derecha en la figura 15.6 se llega a la siguiente política óptima.

Política óptima:
Realizar el sondeo sísmico.
Si el resultado es desfavorable, vender el terreno.
Si el resultado es favorable, perforar en busca de petróleo.
El pago esperado (que incluye los costos del sondeo) es 123 (123 000 dólares).

Es evidente que esta solución óptima (única) es la misma que la que se obtuvo en la sección anterior sin el beneficio de un árbol de decisión. (Vea la política óptima con experimentación que se presenta en la tabla 15.5 y la conclusión, al final de la sección 15.3, que indica que vale la pena realizar la experimentación.)

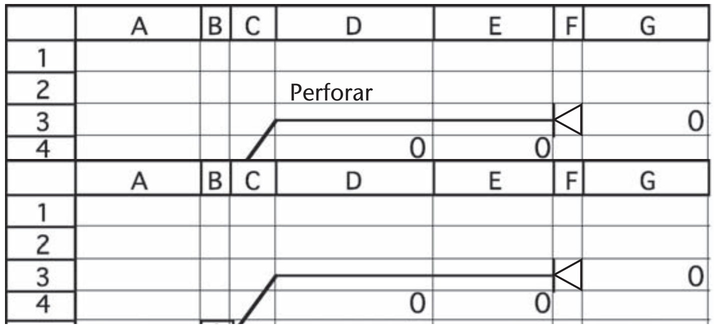
Para cualquier árbol de decisión, este **procedimiento de inducción hacia atrás** siempre conducirá a la *política óptima* (o políticas óptimas) después de calcular las probabilidades de las ramas que salen de un nodo de probabilidad.

En la sección Worked Examples del sitio en internet de este libro se incluye otro ejemplo en el que se resuelve un árbol de decisión de esta manera.

15.5 UTILIZACIÓN DE HOJAS DE CÁLCULO PARA REALIZAR ANÁLISIS DE SENSIBILIDAD EN ÁRBOLES DE DECISIÓN

En la actualidad existen algunos paquetes del tipo hoja de cálculo que son útiles para construir y analizar árboles de decisión. Un complemento de Excel de esta clase que se utiliza de manera extensa se llama *TreePlan*, un software de uso libre desarrollado por el profesor Michael Middleton. La versión académica de TreePlan se incluye en el OR Courseware (junto con su documentación), así como el paquete de uso libre del profesor Middleton SensIt. (Si desea continuar con el uso de alguno de estos paquetes después del curso, será necesario registrarse y pagar los derechos de uso.) Como se mencionó al final de la sección 15.2, SensIt está diseñado para realizar análisis de sensibilidad.

Antes de estudiar SensIt se describirá cómo se utiliza TreePlan para crear un árbol de decisión. Para simplificar esta explicación se comenzará con la ilustración del proceso de construcción de un pequeño árbol de decisión para el primer problema de Goferbroke Co. (sin la consideración de un sondeo sísmico) antes de analizar el problema completo.



■ **FIGURA 15.7**
Árbol de decisión predeterminado creado por TreePlan al seleccionar Decision Tree del menú de herramientas, hacer clic sobre New Tree, para después introducir las etiquetas de Perforar y Vender en las dos alternativas de decisión.

Cómo construye TreePlan el árbol de decisión para el primer problema de la Goferbroke Co.

Considere el primer problema de la Goferbroke Co. (sin sondeo sísmico) como se resumió en la tabla 15.2. Para comenzar a crear el árbol de decisión por medio de TreePlan —después de haber instalado este complemento de Excel[1]—, se selecciona "Decision Tree" de la pestaña Add-Ins (en el caso de Excel 2007) o del menú "Tools" (en el caso de versiones anteriores de Excel), y se da clic sobre "New Tree". Esto crea el árbol de decisión predeterminado que se muestra en la figura 15.7 con un solo nodo de decisión (cuadrado) con dos ramas, que es exactamente lo que se requiere para el primer nodo del problema actual. Sin embargo, aun si se necesitara algo más, en TreePlan resulta fácil hacer cambios a un nodo. Sólo se debe seleccionar la celda que contiene al nodo (B5 en la figura 15.7) y se elige "Decision Tree" del menú "Tools". Ello despliega un cuadro de diálogo que permite cambiar el tipo de nodo —por ejemplo, de un nodo de decisión a un nodo de probabilidad— o agregar más ramas.

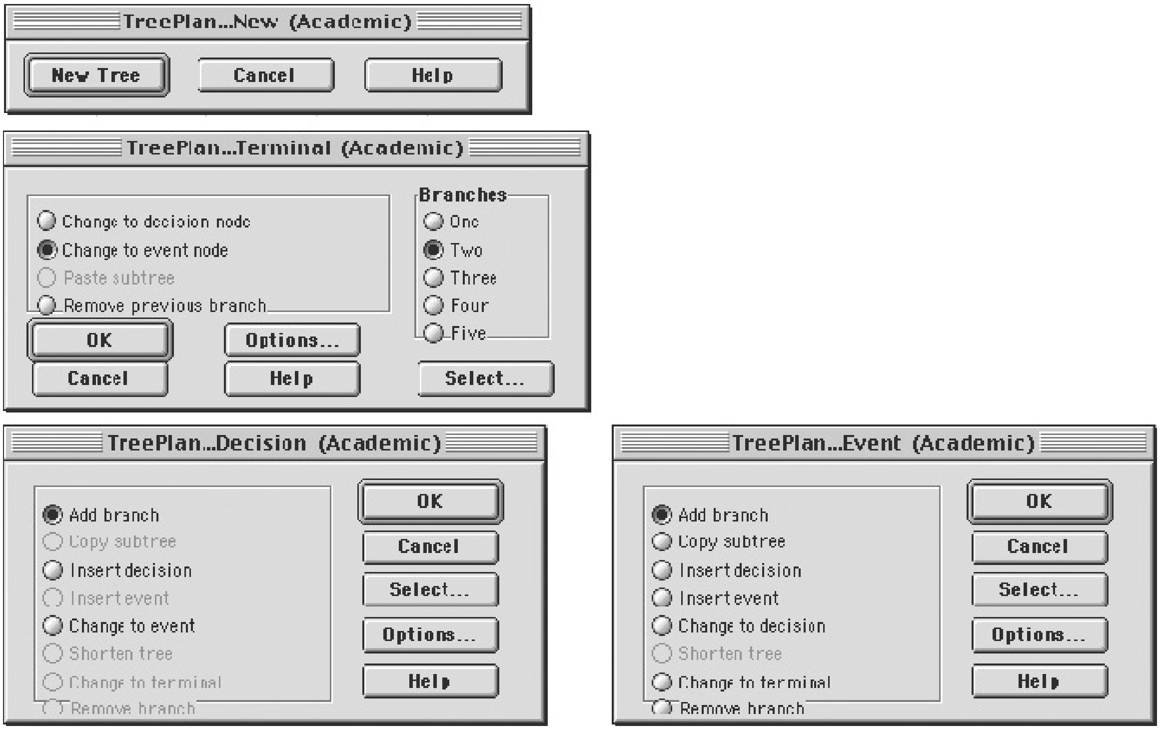
En forma predeterminada, las etiquetas de las decisiones (celdas D2 y D7 en la figura 15.7) son "Decision 1", "Decision 2", etc. Estas celdas se cambian al hacer clic sobre ellas y después digitar una etiqueta nueva. En la figura 15.7, estas etiquetas ya se han cambiado respectivamente a "Perforar" y "Vender".

Si la decisión es perforar, el siguiente evento es saber si el terreno contiene petróleo o no. Para crear un nodo de probabilidad, se da clic sobre la celda que contiene el nodo terminal al final de la rama de perforar (celda F3 en la figura 15.7), y se elige "Decision Tree" de la pestaña Add-Ins o del menú "Tools", luego de lo cual aparecerá el cuadro de diálogo de "TreePlan Acad.-Terminal Node" que se muestra en segundo lugar con respecto a la parte superior de la figura 15.8. Se elige la opción "Change to event node" a la izquierda y se selecciona la opción de dos ramas a la derecha y se hace clic sobre OK. Lo anterior resulta en el árbol de decisión con los nodos y ramas que se muestran en la figura 15.9 (después de reemplazar las etiquetas predeterminadas "Event 1" y "Event 2" con los títulos respectivos de "Petróleo" y "Seco").

En cualquier momento se puede hacer clic en un nodo de decisión existente (cuadrado) o en uno de probabilidad (círculo) y seleccionar Decision Tree del menú Tools para que aparezca el cuadro de diálogo correspondiente —"Tree Plan Acad.-Decision Node" o "TreePlan Acad.-Event Node"— para hacer cualquiera de las modificaciones dadas en la figura 15.8 en ese nodo.

De inicio, cada rama debe mostrar un valor predeterminado de 0 para el flujo de efectivo neto generado en ella (los números aparecen bajo las etiquetas de las ramas: D6, D14, H4 y H9 en la figura 15.9). También, cada una de las dos ramas que salen del evento de probabilidad debe desplegar valores predeterminados de 0.5 para sus probabilidades *a priori* (las probabilidades están justo debajo de las etiquetas correspondientes: H1 y H6 en la figura 15.9). Por lo tanto, en seguida se debe hacer clic sobre estos valores predeterminados y reemplazarlos con los números correctos, a saber,

D6 = –100 (el costo de perforar es 100 000 dólares),
D14 = 90 (la ganancia por la venta es 90 000 dólares),
H1 = 0.25 (la probabilidad *a priori* de petróleo es 0.25),
H4 = 800 (la ganancia neta después de encontrar petróleo es 800 000 dólares),



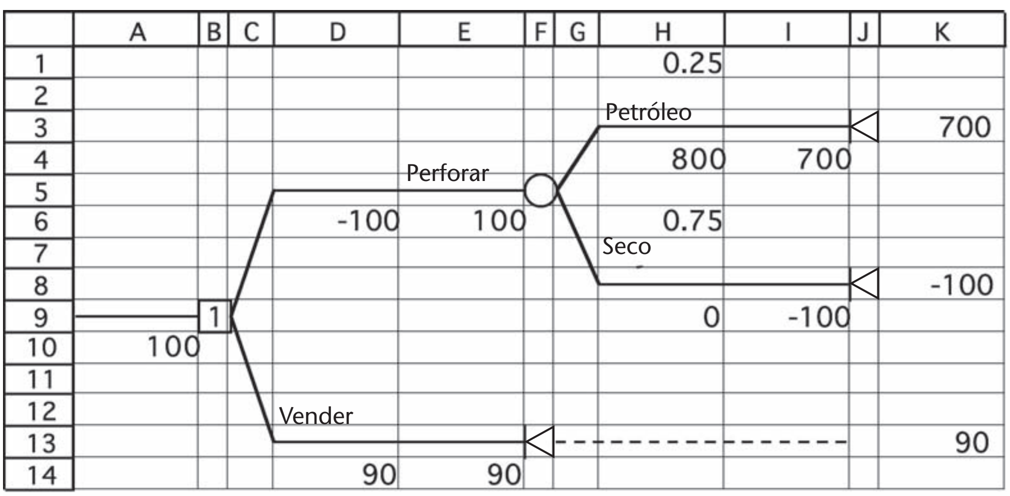
■ **FIGURA 15.8**
Cuadros de diálogo usados por TreePlan para construir un árbol de decisión.

H6 = 0.75 (la probabilidad *a priori* de seco es 0.75),
H9 = 0 (la ganancia neta después de encontrar seco es 0),

como se muestra en la figura.

En cada etapa de la construcción de un árbol de decisión, TreePlan resuelve de manera automática la política óptima con el árbol actual mediante el empleo de la *regla de decisión de Bayes*. El número dentro de cada nodo de decisión indica cuál rama se debería elegir (se supone que las ramas que surgen de ese nodo están numeradas en forma consecutiva de arriba hacia abajo). De esta forma, en el árbol de decisión final de la figura 15.9, el número dentro de la celda B9 especifica que se debería escoger la primera rama (la alternativa de perforar). Los números a ambos lados de

■ **FIGURA 15.9**
Árbol de decisión construido y resuelto por TreePlan para el primer problema de la Goferbroke Co. como se presentó en la tabla 15.2, donde el 1 en la celda B9 indica que se debe elegir la rama superior (la alternativa de perforar).



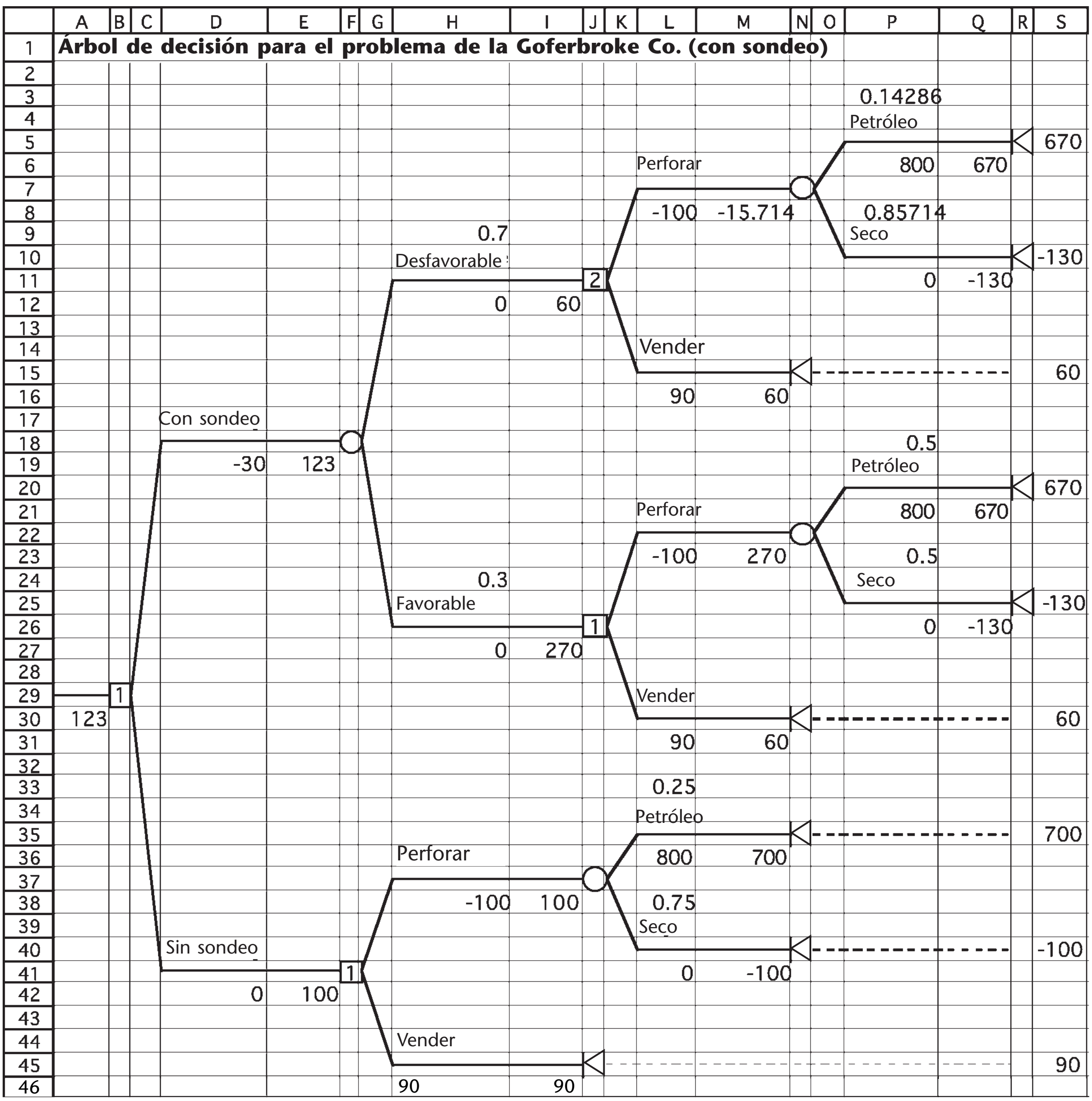
cada nodo terminal significan el pago si se llega a ese nodo. El número 100 en las celdas A10 y E6 es el *pago esperado* en estas etapas del proceso.

Al ejecutar este procedimiento en una computadora, el lector podrá notar que el uso de TreePlan es bastante intuitivo. Si invierte suficiente tiempo en el paquete, también encontrará que éste tiene muchas características útiles que no se han descrito en esta breve introducción.

Árbol de decisión para el problema completo de la Goferbroke Co.

Ahora considere el problema completo de la Goferbroke Co., donde la primera decisión que debe tomarse es si realizar el sondeo sísmico. Continuando con el procedimiento descrito antes, se usa TreePlan para construir y resolver el árbol de decisión que se muestra en la figura 15.10. Aunque

FIGURA 15.10
Árbol de decisión construido y resuelto por TreePlan para el problema completo de la Goferbroke Co. que también considera la opción de realizar o no un sondeo sísmico.



la forma es algo diferente, observe que este árbol de decisión es equivalente por completo al de la figura 15.6. Además de la conveniencia de construir el árbol de decisión de manera directa en una hoja de cálculo, TreePlan también proporciona la ventaja clave de resolver de manera automática el árbol de decisión. En lugar de confiar en cálculos hechos a mano como en la figura 15.6, TreePlan calcula en forma instantánea todos los pagos esperados en cada etapa del árbol, como se muestra en seguida de cada nodo, tan pronto como se construye el árbol de decisión. En lugar de usar diagonales dobles, TreePlan pone un número dentro de cada nodo de decisión con el que indica cuál rama se debe escoger (se supone que las ramas que surgen de cada nodo están numeradas en forma consecutiva de arriba hacia abajo).

Organización de la hoja de cálculo para realizar análisis de sensibilidad

Al final de la sección 15.2 se explicó cómo se pueden realizar análisis de sensibilidad en un pequeño problema (el primero de la Goferbroke Co.), en el que sólo se necesitaba tomar una decisión (perforar o vender). En ese caso, el análisis fue bastante directo porque el pago esperado de cada alternativa de decisión se podría expresar como una función simple del parámetro (la probabilidad *a priori* de petróleo) del modelo considerado. En contraste, cuando debe tomarse una secuencia de decisiones, como para el problema completo de la Goferbroke Co., el análisis de sensibilidad adquiere mayor importancia. Ahora en el modelo hay más parámetros (los diferentes costos, ganancias y probabilidades) lo que podría significar suficiente incertidumbre como para garantizar la realización del análisis de sensibilidad. Aún más, para encontrar el pago máximo esperado de cualesquiera valores particulares de los parámetros del modelo, ahora se debe resolver un árbol de decisión. Por lo tanto, la utilización de un paquete de hoja de cálculo como TreePlan que resuelve de manera automática el árbol de decisión se vuelve muy útil. Al agregar software que está diseñado específicamente para realizar análisis de sensibilidad, como SensIt, pronto se obtienen otras perspectivas.

Si se inicia con una hoja de cálculo que ya contiene un árbol de decisión, el siguiente paso es expandirla y organizarla para realizar el análisis de sensibilidad. Ahora se ilustrará este caso para el problema completo de la Goferbroke Co. a partir de la hoja de cálculo de la figura 15.10 que contiene el árbol de decisión que construyó TreePlan.

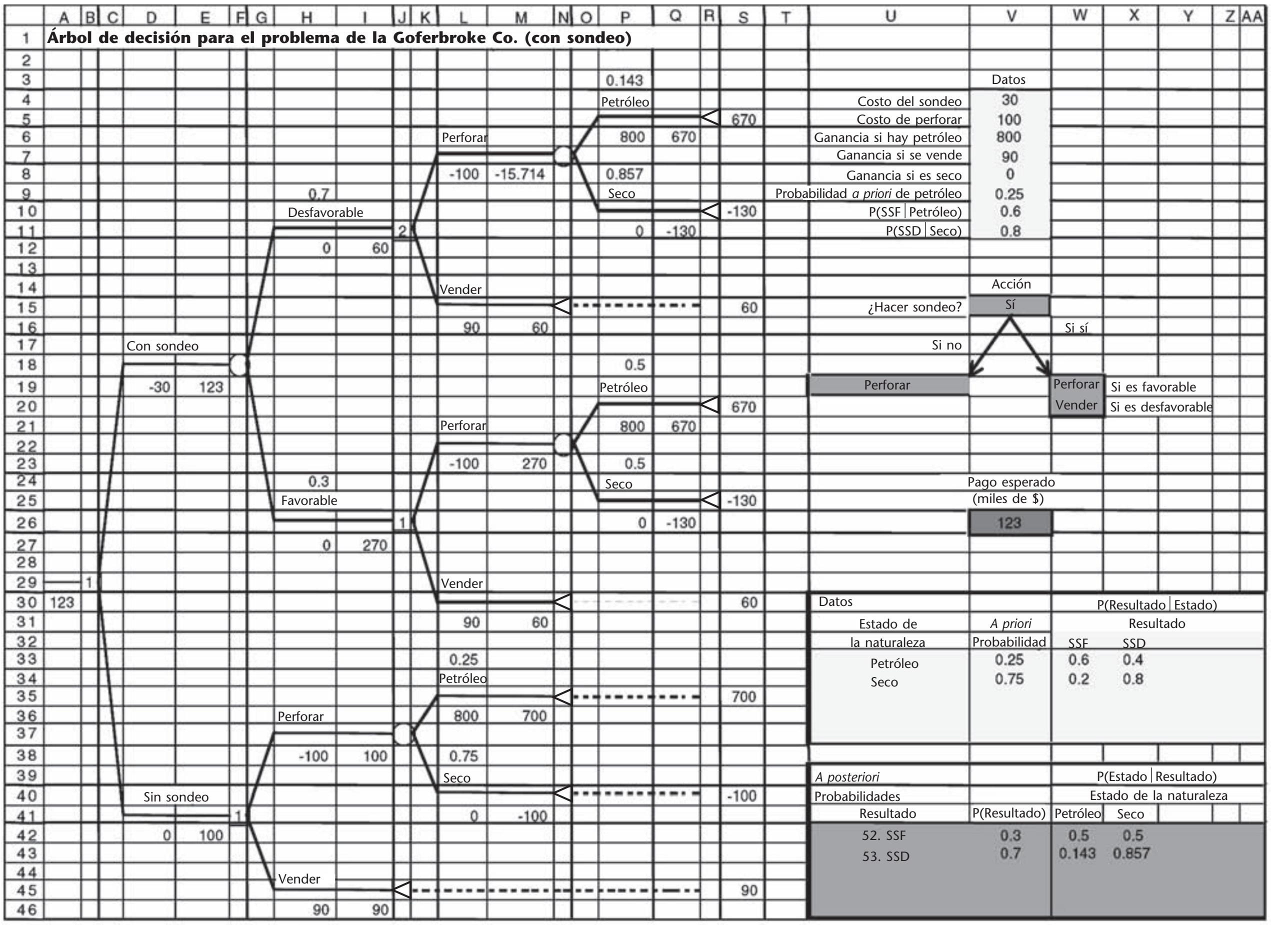
Resulta útil comenzar por consolidar los datos y resultados en una nueva sección, como se muestra en el lado derecho de la figura 15.11. Todas las celdas de datos del árbol de decisión ahora necesitarían hacer referencia a las celdas de datos consolidados (celdas V4:V11), como lo ilustran las fórmulas que se presentaron en las celdas P6 y P11 al final de la figura. En forma similar, los resultados resumidos a la derecha del árbol de decisión hacen referencia a las celdas de salida dentro del árbol de decisión —los nodos de decisión en las celdas B29, F41, J11 y J26, así como el pago esperado en la celda A30— al usar las fórmulas de las celdas U19, V15, V26, y W19:W20 que se despliegan al final de la figura 15.11.

Los datos de probabilidad del árbol de decisión están complicados por el hecho de que las probabilidades *a posteriori* necesitarán actualizarse cada vez que se haga un cambio en cualquiera de los datos de las probabilidades *a priori*. Por fortuna, la plantilla para calcular probabilidades *a posteriori* (como se muestra en la figura 15.3) se puede utilizar para realizar estos cálculos. La parte relevante de esta plantilla (B3:H19) se ha copiado (mediante los comandos "Copiar" y "Pegar" del menú "Edición") a la hoja de cálculo de la figura 15.11 (ahora aparece en U30:AA46). Los datos de la plantilla se refieren a los datos de probabilidad en las celdas de datos ProbabilidadApriori-DePetroleo (V9), ProbSSFDadoPetroleo (V10), y ProbSSDDadoSeco (V11), como se muestra en las fórmulas de las celdas V33:X34 al final de la figura 15.11. La plantilla calcula de manera automática la probabilidad de cada resultado y las probabilidades *a posteriori* (en las celdas V42:X43) basadas en estos datos. El árbol de decisión acude después a estas probabilidades calculadas cuando las mismas sean necesarias, como se muestra en las fórmulas de las celdas P3:P11 en la figura 15.11.

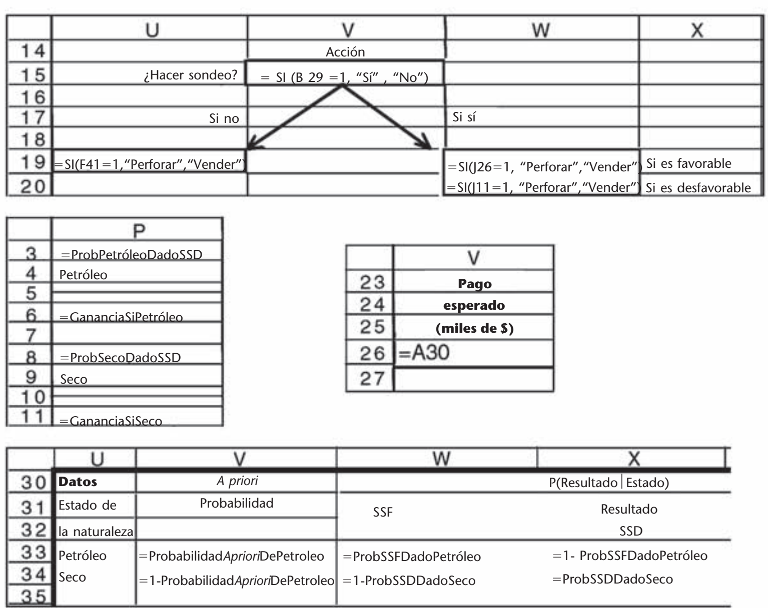
La consolidación de datos y resultados ofrece dos ventajas. Primero, asegura que cada dato está sólo en un lugar. Cada vez que ese dato se necesita en el árbol de decisión, se hace una referencia a una sola celda de datos, lo cual simplifica en gran medida el análisis de sensibilidad. Para cambiar un dato, es necesario cambiarlo sólo en un lugar en vez de buscar a través de todo el árbol para encontrar y cambiar todas las ocurrencias de ese dato. Una segunda ventaja de la consolidación de datos y resultados es que permite que *cualquiera* interprete el modelo. No es necesario entender

FIGURA 15.11

Como preparación para realizar análisis de sensibilidad sobre el problema completo de la Goferbroke Co., los datos y resultados se han consolidado en la hoja de cálculo a la derecha del árbol de decisión.



Nombre de Rango	Celda
CostoDePerforar	V5
CostoDeSondeo	V4
Probabilidad*Apriori*DePetróleo	V9
ProbabilidadSecoDadoSSF	X42
ProbabilidadSecoDadoSSD	X43
ProbSSF	V42
ProbSSFDadoPetróleo	V10
ProbPetróleoDadoSSF	W42
ProbPetróleoDadoSSD	W43
ProbSSD	V43
ProbSSDDadoSeco	V11
GananciaSiSeco	V8
GananciaSiPetróleo	V6
GananciaSiVende	V7



TreePlan o cómo leer un árbol de decisión para ver cuáles datos se usaron en el modelo o cuáles son el plan de acción sugerido y el pago esperado.

Aunque se debe invertir algo de tiempo y esfuerzo para consolidar los datos y resultados, entre ellos todas las referencias cruzadas necesarias, este paso es verdaderamente esencial para realizar el análisis de sensibilidad. Muchos datos se utilizan en varias partes del árbol de decisión. Por ejemplo, la ganancia si Goferbroke encuentra petróleo aparece en las celdas P6, P21 y L36. Para realizar análisis de sensibilidad sobre este dato ahora se debe cambiar su valor sólo en un sitio (celda V6) en vez de hacerlo en tres (celdas P6, P21 y L36). Los beneficios de la consolidación son aún más importantes para los datos de probabilidad. Un cambio en cualquier probabilidad *a priori* puede causar que cambien *todas* las probabilidades *a posteriori*. Mediante la inclusión de la plantilla de probabilidad *a posteriori* se puede cambiar la probabilidad *a priori* en un sitio, y entonces todas las otras probabilidades se calculan y actualizan de la manera apropiada.

Después de hacer cualquier cambio en los datos de costo, ganancia o probabilidad de la figura 15.11, la hoja de cálculo resume de manera ágil los nuevos resultados, una vez que el trabajo real para obtener esta información se realiza en forma instantánea por la plantilla de probabilidades *a posteriori* y el árbol de decisión. Por lo tanto, la experimentación con datos alternativos mediante prueba y error es una forma útil de realizar análisis de sensibilidad.

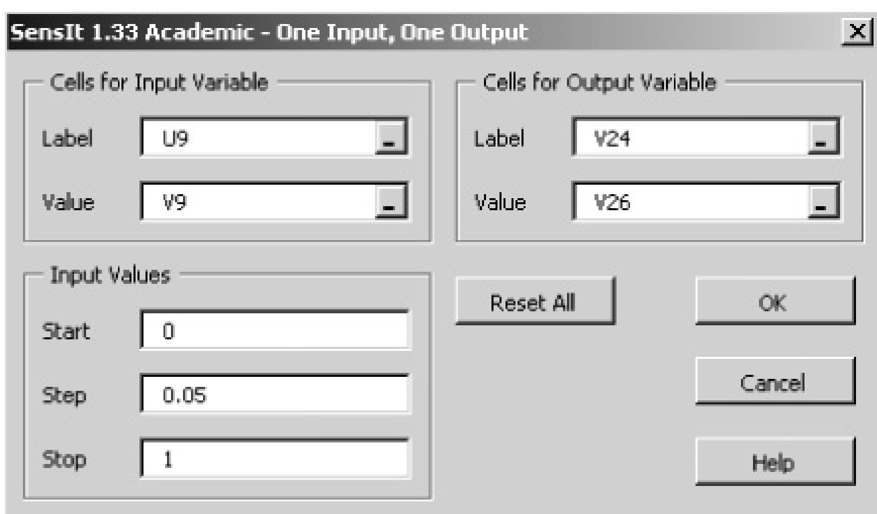
Sin embargo, sería deseable tener otro método para realizar análisis de sensibilidad en forma más sistemática. Aquí es donde SensIt adquiere mayor utilidad. Proporciona un modo fácil de crear en forma sistemática gráficas informativas del análisis de sensibilidad que despliegan el efecto de cambiar un número en las celdas de datos que se desee. SensIt está diseñado para integrarse con TreePlan (aunque también puede realizar otros tipos de análisis de sensibilidad que no requieren el uso de TreePlan).

Utilización de SensIt para crear tres tipos de gráficas de análisis de sensibilidad

Al instalar SensIt aparece un comando adicional de Análisis de Sensibilidad en la pestaña Add-Ins (en el caso de Excel 2007) o el menú de herramientas (en el caso de versiones anteriores de Excel). Este elemento tiene un submenú que proporciona las opciones de dos diferentes tipos de gráficas de análisis de sensibilidad: 1) dibujar una gráfica de una *sola salida* (como, por ejemplo, la carga esperada) *versus* una *sola entrada* (como, por ejemplo, la probabilidad *a priori* de encontrar petróleo) o 2) generar gráficas que, de manera simultánea, comparen el efecto de *entradas múltiples* en una *sola salida*. A continuación se describirán estos dos tipos de análisis de sensibilidad.

Seleccionar la opción de dibujar una gráfica de una sola salida *versus* una sola entrada provoca que aparezca el cuadro de diálogo que se muestra en la figura 15.12. La mitad superior de ese cuadro de diálogo se usa para especificar la celda de datos que se cambiará (la probabilidad *a priori*

FIGURA 15.12
Cuadro de diálogo usado por la opción SensIt para dibujar una gráfica de una sola salida *versus* una sola entrada.



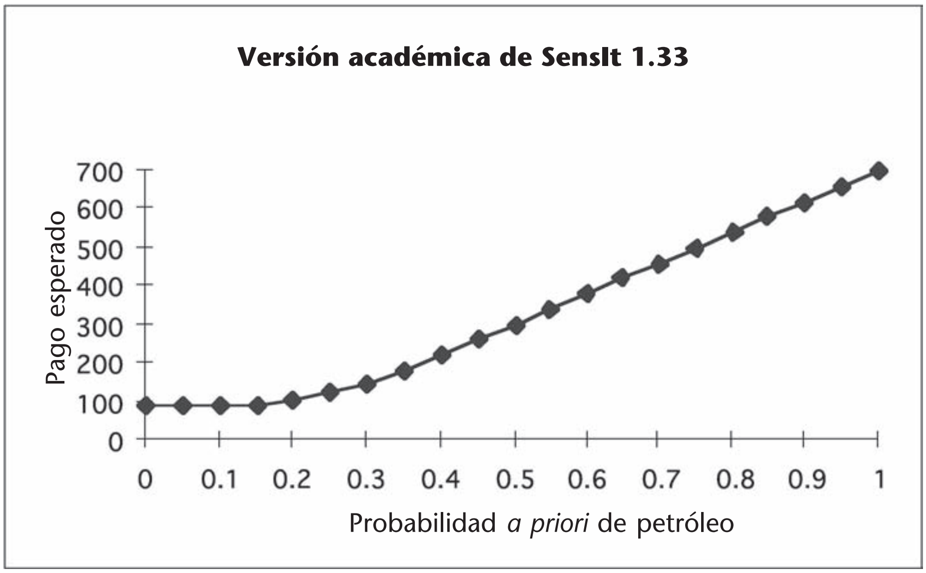


FIGURA 15.13
Gráfica generada por la opción Plot de SensIt para el problema completo de la Goferbroke Co. para mostrar cómo el pago esperado (cuando se usa la regla de decisión de Bayes) depende de la probabilidad *a priori* de petróleo.

de encontrar petróleo en la celda V9) y la celda de salida de interés (el pago esperado en la celda V26). De manera opcional, también se pueden especificar las celdas que contienen las etiquetas de estas celdas (celdas U9 y V24, respectivamente). Estas etiquetas se usan para identificar los ejes de la gráfica que se crea. La mitad inferior del cuadro de diálogo se utiliza para especificar el rango de valores que será considerado para una sola celda de datos (la probabilidad *a priori* de petróleo). En este caso se considerarán todos los valores entre 0 y 1 (en intervalos de 0.05). Cuando se hace clic sobre OK se genera la gráfica que se muestra en la figura 15.13, la cual revela la relación entre la probabilidad *a priori* de encontrar petróleo y el pago esperado que resulta cuando se aplica la política óptima dada esta probabilidad.

La gráfica indica que el pago esperado comienza a crecer cuando la probabilidad *a priori* es un poco mayor que 0.15 y después crece con más rapidez cuando esta probabilidad está alrededor de 0.3. Ello sugiere que la política óptima cambia en puntos cercanos a estos valores de la probabilidad *a priori*. Para verificar este cambio se puede utilizar la hoja de cálculo de la figura 15.11 para ver cómo cambian los resultados cuando la probabilidad *a priori* de encontrar petróleo se incrementa con lentitud en la vecindad de estos valores. Este tipo de análisis de prueba y error pronto conduce a las siguientes conclusiones acerca de cómo la política óptima depende de esta probabilidad.

Política óptima

Sea p = Probabilidad *a priori* de encontrar petróleo.
Si $p \leq 0.168$, se vende el terreno (sin sondeo sísmico).
Si $0.169 \leq p \leq 0.308$, se realiza el sondeo sísmico: se perfora si éste es favorable y se vende en caso contrario.
Si $p \geq 0.309$ se perfora en busca de petróleo (sin sondeo sísmico).

Hasta ahora, este análisis de sensibilidad se ha enfocado en investigar el efecto si la probabilidad verdadera de encontrar petróleo es diferente de la probabilidad *a priori* de 0.25. Se puede realizar un análisis similar con respecto a las probabilidades de las celdas V10:V11 de la figura 15.11. Sin embargo, como existe una incertidumbre significativa acerca de los datos de costo y ganancia de las celdas V4:V7, en seguida se realizará análisis de sensibilidad con respecto a estos últimos datos.

Suponga que se desea investigar cómo podría cambiar el pago esperado si uno de los costos o beneficios de las celdas V4:V7 cambiara. Esta nueva situación requiere hacer algunas adiciones a la hoja de cálculo original (figura 5.11). Como se muestra en la figura 15.14 se agregan tres columnas por cada celda de datos que se modificará, lo cual indica el valor más bajo, el valor básico y el valor más alto. Suponga que el costo del estudio y la ganancia en el caso de que se vendiera el

■ **FIGURA 15.14**
Vista ampliada de la hoja de cálculo de la figura 15.11 para prepararse a usar SensIt para investigar el efecto de un cambio de valor en cualquier costo o ganancia en el pago esperado.

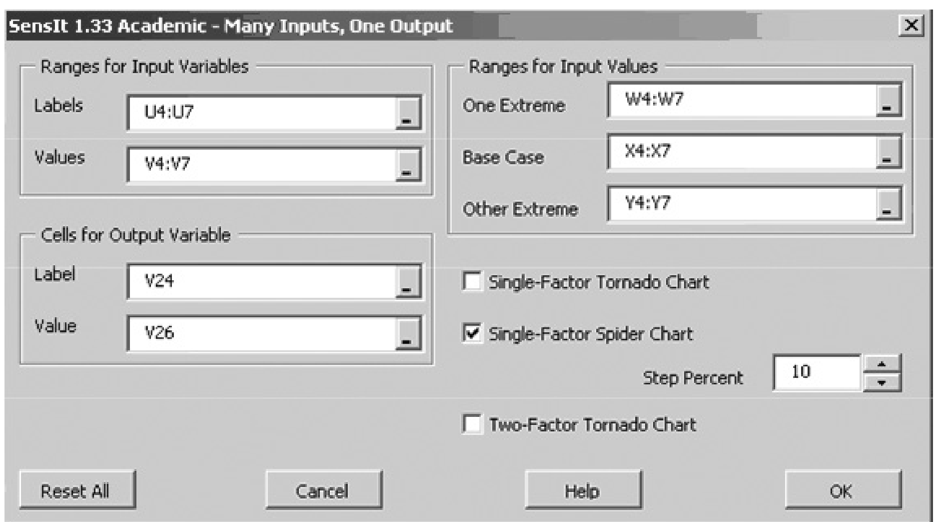
	U	V	W	X	Y
3		Data	Low	Base	High
4	Cost of Survey	30	28	30	32
5	Cost of Drilling	100	75	100	140
6	Revenue if Oil	800	600	800	1000
7	Revenue if Sell	90	85	90	95
8	Revenue if Dry	0			
9	Prior Probability of Oil	0.25			
10	P(FSS|Oil)	0.6			
11	P(USS|Dry)	0.8			

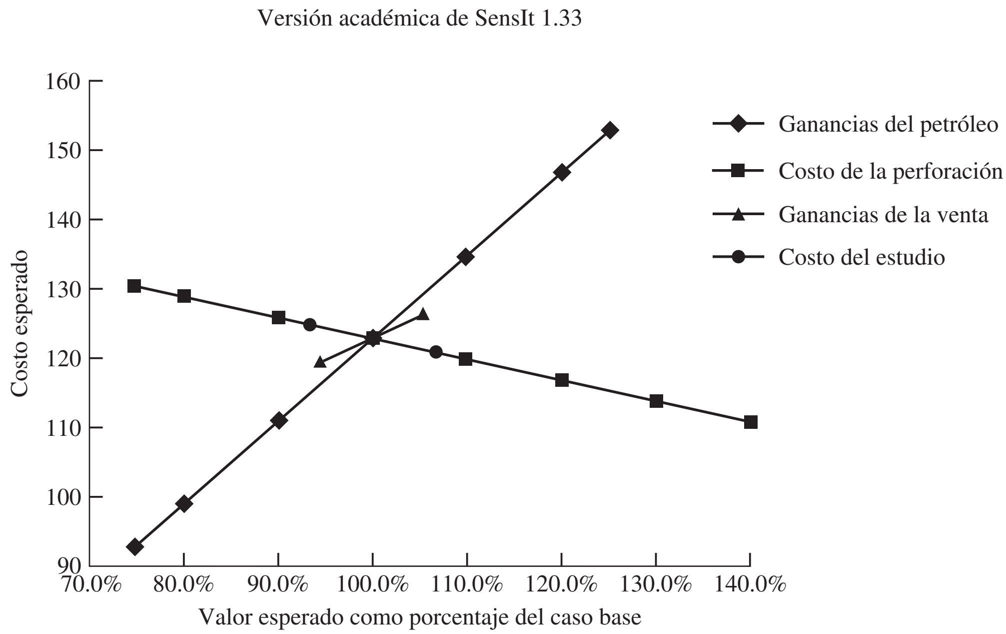
terreno fueran muy predecibles (es decir, que variara en un rango pequeño, digamos de 28 a 32 y de 85 a 95, respectivamente), mientras que el costo de perforación y la ganancia si el petróleo es extraído son más variables (es decir, que varían en un rango grande, digamos de 75 a 140 y de 600 a 1 000, respectivamente).

Ya que deseamos investigar cómo cambiaría el costo esperado si cualquiera de los costos o ganancias incluidos en las celdas V4:V7 cambiara, ahora tenemos cuatro entradas (dichos costos y ganancias) y una salida (el costo esperado). Por lo tanto, después de expandir la hoja de cálculo como se muestra en la figura 15.14, el siguiente paso es acceder al cuadro de diálogo SensIt “many inputs, one output”. Esta caja de diálogo (a la cual se accede seleccionando al artículo correspondiente en el menú Análisis de sensibilidad en la pestaña Add-Ins de Excel 2007 o en el menú Tools en el caso de versiones anteriores de Excel) se muestra en la figura 15.15. Se utiliza para especificar qué celdas de datos contiguas se modificarán, qué celda de salida se analizará y la ubicación de las celdas que especifican el rango (abajo, en la base o arriba) de las celdas de datos. La opción Step Percent se utiliza para especificar el tamaño deseado del paso (como porcentaje del valor base) en cada valor de entrada al cual el costo esperado se recalculará hasta que se llegue a los valores bajo y alto de la entrada. En el extremo inferior derecho de la caja de diálogo se ofrece la opción de tener tres gráficas para desplegar, el efecto que tendrá en la salida y que tendrá tener valores alternados en cualquiera de estas entradas. Suponga que se selecciona la opción “single-factor spider chart” (como se muestra en la figura 15.16). Al seleccionar la opción OK se genera la **gráfica spider** que se muestra en la figura 15.17.

Cada línea de la gráfica Spider de esta figura representa el pago esperado cuando se cambia una de las celdas de datos seleccionadas (V4:V7) de su valor original a otro que resulta de multiplicar éste por el porcentaje que se indica en la parte baja de la gráfica. (El *costo de la línea del estudio* se encuentra en la parte superior de la *línea del costo de perforación*.) Sin embargo, es mucho más corta que la última línea puesto que ésta sólo se extiende 93.3% hacia su lado izquierdo y 106.7%

■ **FIGURA 15.15**
Cuadro de diálogo usado por SensIt para investigar de manera simultánea el efecto que tiene el cambio en cualquiera de las entradas sobre la salida.



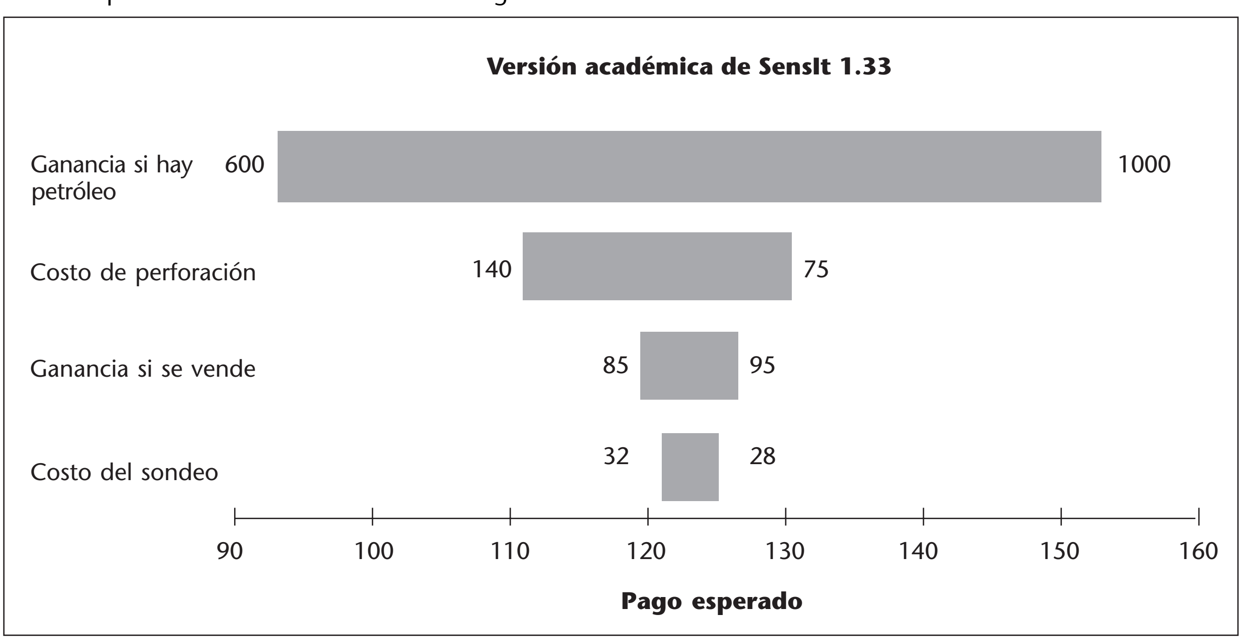


■ **FIGURA 15.16**
Gráfica spider generada por SensIt para el problema Goferbroke completo que muestra cómo varía el costo esperado (cuando se usa la regla de decisión de Bayes) al cambiar cualquiera de las estimaciones de ganancia o costo.

hacia el derecho). El hecho de que la *recta de la ganancia si hay petróleo* sea la más inclinada revela que el pago esperado es particularmente sensible a la estimación de la ganancia si se encuentra petróleo, por lo que cualquier trabajo adicional sobre el refinamiento de esta estimación debe recibir la mayor atención posible.

Ahora suponga que la opción "single-factor tornado chart" de la figura 15.15 se selecciona. Si se presiona la opción OK se genera la **gráfica de tornado** que se muestra en la figura 15.17. Cada barra en la gráfica muestra el rango de variación del costo esperado a medida que el costo corres-

■ **FIGURA 15.17**
Diagrama de tornado generado por la opción Tornado de SensIt para el problema completo de la Goferbroke Co. para mostrar qué tanto puede variar el pago esperado (cuando se usa la regla de decisión de Bayes) en el intervalo completo de valores probables de cualquiera de las estimaciones de costo o ganancia.



pondiente o ganancia varía en el rango de valores indicado numéricamente en los extremos de cada barra. El ancho de cada barra de la gráfica es una medida del grado de sensibilidad del costo esperado respecto de los cambios en ese costo de la barra o ganancia. Una vez más, la ganancia *si hay petróleo* permanece y causa una sensibilidad mucho mayor que los demás costos o ganancias.

La gráfica spider de la figura 15.16 y la de tornado de la figura 15.17 proporcionan la misma información en formas complementarias. En realidad, es una cuestión de gustos determinar cuál de estas gráficas contiene dicha información de una manera más real.

No se analizará la tercera opción (la "gráfica de tornado con dos factores") de la figura 15.15. En la Guía de usuario que se encuentra en el sitio en internet de este libro se puede consultar información adicional y documentación completa de SensIt (en cuanto respecta a TreePlan).

15.6 TEORÍA DE LA UTILIDAD

Hasta ahora, cuando se aplica la regla de decisión de Bayes, se ha supuesto que el pago esperado en *términos monetarios* es la medida adecuada de las consecuencias de optar por una opción. Sin embargo, en muchas situaciones este supuesto no es apropiado.

Por ejemplo, suponga que se ofrece a un individuo la oportunidad de: 1) aceptar 50% de posibilidades de ganar 100 000 dólares o nada, o 2) recibir 40 000 dólares con seguridad. Muchas personas preferirán los 40 000 dólares aun cuando el pago esperado con 50% de posibilidades de ganar 100 000 dólares es 50 000 dólares. Una compañía no siempre estará dispuesta a invertir una gran suma de dinero en un nuevo producto, aunque la ganancia esperada sea sustanciosa, si existe el riesgo de perder la inversión y quedar en bancarrota. Las personas compran seguros aunque sea una mala inversión desde el punto de vista del pago esperado.

¿Invalidan estos ejemplos la regla de decisión de Bayes? Por fortuna, la respuesta es no, porque los *valores monetarios* se pueden transformar en una escala adecuada que refleje las preferencias del tomador de decisiones. Esta escala se llama *función de utilidad del dinero*.

Función de utilidad del dinero

En la figura 15.18 se muestra una **función de utilidad común U(*M*) de la cantidad de dinero *M***. La figura indica que un individuo que tiene esta función de utilidad valora la obtención de 30 000 dólares en el doble que 10 000 dólares y valoraría la obtención de 100 000 dólares en el doble que obtener 30 000. Esta preferencia refleja el hecho de que las necesidades de más alta prioridad de una persona quedarían satisfechas con los primeros 10 000 dólares. Cuando la pendiente de la función disminuye conforme aumenta la cantidad de dinero se dice que existe una **utilidad marginal decreciente del dinero**. Se dice que este individuo tiene **aversión al riesgo**.

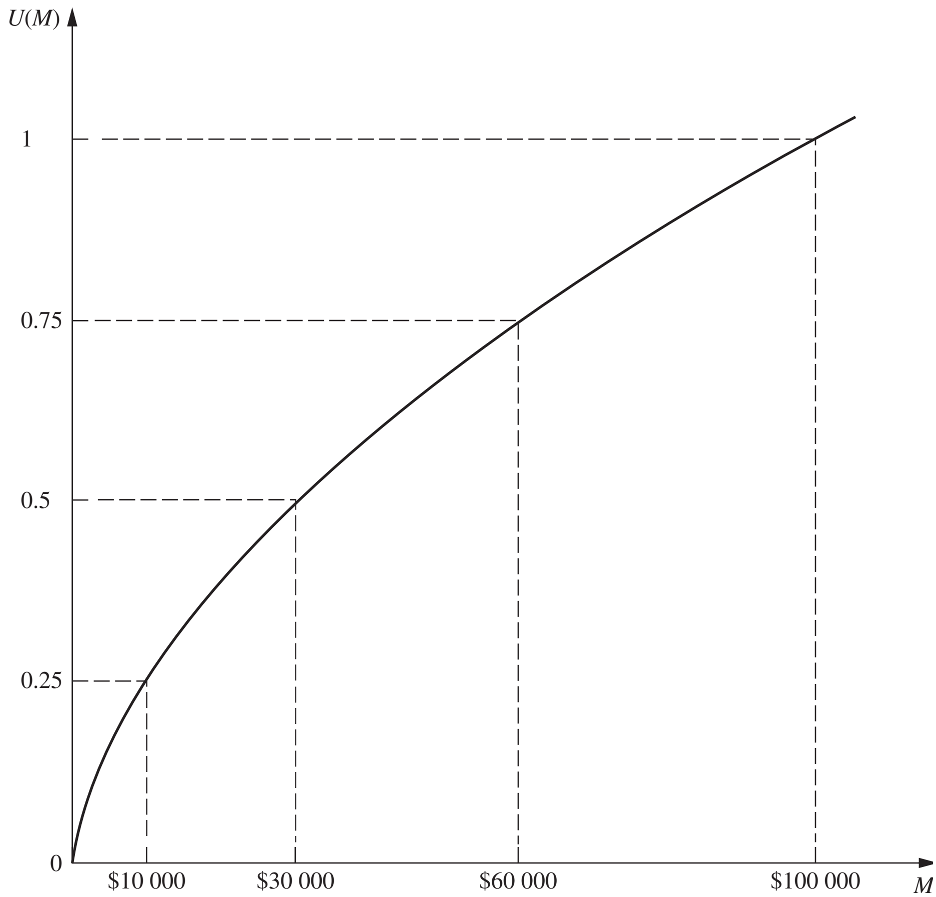
Sin embargo, no todas las personas muestran una utilidad marginal decreciente del dinero. Algunas **buscan el riesgo** en lugar de sentir *aversión al riesgo* y van por la vida buscando el "premio mayor". La pendiente de su función de utilidad *aumenta* a medida que la cantidad de dinero crece, de manera que tienen una **utilidad marginal creciente del dinero**.

El caso intermedio es un individuo **neutral al riesgo**, que aprecia el dinero por lo que vale. La función de utilidad de este individuo es proporcional a la cantidad de dinero involucrada. Aunque algunas personas parecen neutrales al riesgo cuando se trata de pequeñas cantidades, no es usual que sean en verdad neutrales al riesgo con grandes sumas.

También es posible mostrar una mezcla de estos tipos de comportamiento. Por ejemplo, un individuo puede en esencia ser neutral al riesgo con poco dinero, convertirse en un buscador de riesgo con cantidades moderadas y sentir aversión al riesgo ante las grandes sumas. Además, la actitud hacia el riesgo puede cambiar con el tiempo y las circunstancias.

La actitud de un individuo hacia el riesgo si se trata de las finanzas personales puede diferir de cuando se toman decisiones dentro de una organización. Por ejemplo, los administradores de una empresa deben considerar las circunstancias de la compañía y la filosofía colectiva de la alta administración para determinar la actitud adecuada ante el riesgo para tomar esas decisiones gerenciales.[3]

[3] Un sondeo de la forma de la función utilidad de 332 propietarios-gerentes y el impacto de ésta en el comportamiento organizacional, consulte J. M. E. Pennings y A. Smidts, "The Shape of Utility Functions and Organizational Behavior", en *Management Science*, **49**: 1251-1263, 2003.



■ **FIGURA 15.18**
Función de utilidad típica para el dinero, donde *U*(*M*) es la utilidad de obtener una cantidad de dinero *M*.

El hecho de que distintas personas tienen funciones de utilidad del dinero diferentes tiene una aplicación importante para el tomador de decisiones frente a la incertidumbre.

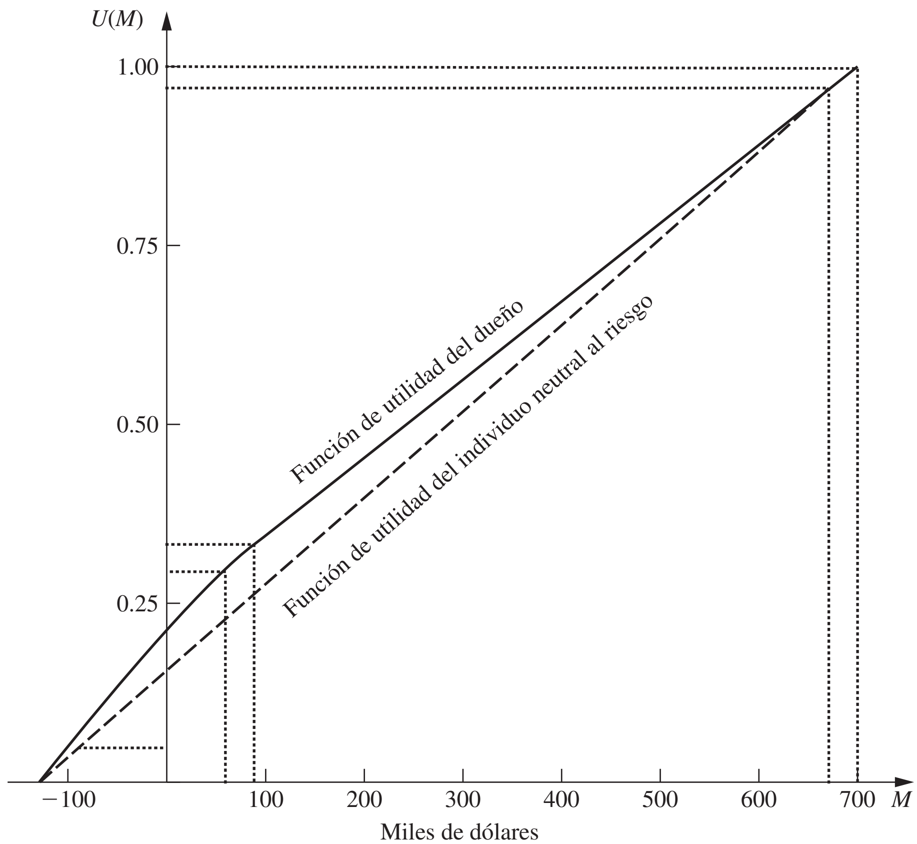
> Cuando la *función de utilidad del dinero* se incorpora en un enfoque de análisis de decisiones con respecto a un problema, esta función de utilidad debe construirse de manera que se ajuste a las preferencias y valores del tomador de decisiones. (Éste puede ser un solo individuo o bien un grupo de personas.)

La *escala* de la función utilidad es irrelevante. En otras palabras, no importa si el valor de $U(M)$ en las líneas discontinuas de la figura 15.18 son 0.25, 0.5, 0.75, 1 (como se muestra) o 10 000, 20 000, 30 000, 40 000 o lo que sea. Todas las utilidades pueden multiplicarse por cualquier constante positiva sin afectar qué curso de acción alterno tendrá la utilidad esperada mayor. También es posible agregar la misma constante (positiva o negativa) a todas las utilidades sin afectar qué curso de acción tendrá la utilidad esperada mayor.

Por esas razones, existe libertad para fijar el valor de $U(M)$ de forma arbitraria para dos valores de M, siempre y cuando el mayor valor monetario tenga la utilidad más elevada. Es particularmente conveniente (aunque no necesario) fijar el valor de $U(M) = 0$ al valor más pequeño de M en consideración y fijar el valor de $U(M) = 1$ al valor más grande de M, como se hizo en la figura 15.18. Mediante la asignación de una utilidad de 0 a la peor salida y una utilidad de 1 a la mejor, para después determinar las utilidades de las demás salidas, es muy fácil ver la utilidad relativa de cada salida a los largo de la escala desde la peor hasta la mejor.

La clave para considerar que la función de utilidad del dinero se ajuste al tomador de decisiones es la siguiente propiedad fundamental de las funciones de utilidad.

> **Propiedad fundamental:** Bajo los supuestos de la teoría de utilidad, la *función de utilidad del dinero* de un tomador de decisiones tiene la propiedad de que éste se muestra



■ **FIGURA 15.19**
Función de utilidad del dinero del dueño de la Goferbroke Co.

indiferente entre dos cursos de acción alternativos si los dos tienen la *misma utilidad esperada*.

Por ejemplo, suponga que el tomador de decisiones tiene la función de utilidad que se muestra en la figura 15.19. Suponga además que se ofrece a este tomador de decisiones la siguiente oportunidad.

Oferta: Obtener \$100 000 (utilidad = 1) con probabilidad p o nada (utilidad = 0) con probabilidad $(1 - p)$.

Entonces,

$E(\text{utilidad}) = p,$ para esta oferta.

Por lo tanto, en el caso de *cada* uno de los siguientes tres pares de alternativas, el tomador de decisiones es indiferente entre la primera y la segunda:

1. La oferta con $p = 0.25$ [$E(\text{utilidad}) = 0.25$] o definitivamente obtener \$10 000 (utilidad = 0.25)
2. La oferta con $p = 0.5$ [$E(\text{utilidad}) = 0.5$] o definitivamente obtener \$30 000 (utilidad = 0.5)
3. La oferta con $p = 0.75$ [$E(\text{utilidad}) = 0.75$] o definitivamente obtener \$60 000 (utilidad = 0.75)

Este ejemplo ilustra también la forma en que se puede construir la función de utilidad del dinero del tomador de decisiones desde el principio. Se haría al tomador de decisiones la misma oferta hipotética de obtener una gran suma de dinero (por ejemplo, 100 000 dólares) con probabilidad p u obtener nada. Después, para cada una de las cantidades pequeñas (10 000, 30 000 y 60 000 dólares), se pediría al tomador de decisiones que eligiera un valor de p ante el que se sintiera

indiferente entre la oferta y la obtención definitiva o segura de esa cantidad de dinero. Entonces, la utilidad de una cantidad pequeña de dinero es p veces la de utilidad de la cantidad grande. Este procedimiento, llamado el *método de la lotería equivalente* para determinar las utilidades se describe a continuación.

Método de la lotería equivalente

1. Determine el mayor costo potencial, $M = maximum$ y asígnelo a alguna utilidad, por ejemplo, $U(maximum) = 1$.
2. Determine el menor costo potencial, $M = minimum$ y asígnelo a alguna utilidad más pequeña que la del paso 1, por ejemplo, $U(minimum) = 0$.
3. Para determinar la utilidad de otro costo potencial M, al tomador de decisiones se le ofrecen las dos alternativas hipotéticas siguientes:

   A1: Obtener un costo de *maximum* con una probabilidad p,
   Obtener un costo de *minimum* con una probabilidad $1 - p$

   A2. Definitivamente obtener un costo de M.

   Pregunta para el tomador de decisiones: ¿Qué valor de p lo hace a usted *indiferente* entre estas dos alternativas? La utilidad resultante M es, entonces,

$$U(M) = p\, U(maximum) + (1 - p)\, U(minimum),$$

   que se puede simplificar en

$$U(M) = p, \qquad \text{si } U(minimum) = 0, \qquad U(maximum) = 1.$$

Ahora se puede resumir el papel básico de las funciones de utilidad en un análisis de decisiones.

> Cuando la función de utilidad del dinero del tomador de decisiones se usa para medir el valor relativo de los distintos resultados monetarios posibles, la *regla de decisión de Bayes* sustituye los pagos monetarios por las utilidades correspondientes. Por lo tanto, la acción óptima (o serie de acciones óptimas) es la que *maximiza la utilidad esperada*.

Sólo se estudiaron aquí las funciones de utilidad *del dinero*. No obstante, debe mencionarse que en ocasiones pueden construirse funciones de utilidad cuando algunas o todas las consecuencias de los diferentes cursos de acción *no* son monetarias. (Por ejemplo, las consecuencias de las alternativas de decisión de un médico al tratar a un paciente involucran la salud futura de éste.) Esto no necesariamente es sencillo, ya que puede requerirse hacer juicios de valor sobre qué tan deseables, de manera relativa, son algunas consecuencias más o menos intangibles. De todas formas, en esas circunstancias, es importante incorporar esos juicios de valor al proceso de decisión.

Aplicación de la teoría de utilidad al problema completo de Goferbroke Co.

Al final de la sección 15.1 se mencionó que Goferbroke Co. operaba con poco capital, por lo que una pérdida de 100 000 dólares sería bastante seria. El dueño (mayoritario) de la compañía ha adquirido una gran deuda para continuar con la operación. El peor escenario sería conseguir 30 000 dólares para un sondeo sísmico y después todavía perder 100 000 dólares en la perforación cuando no hay petróleo. Esta situación no llevaría a la compañía a la bancarrota por ahora, pero es claro que la dejaría en una posición financiera precaria.

Por otro lado, encontrar petróleo es una perspectiva interesante, puesto que una ganancia de 700 000 dólares daría a la compañía, por fin, una base financiera sólida.

Para aplicar la *función de utilidad del dinero* del dueño (el tomador de decisiones) al problema descrito en las secciones 15.1 y l5.3 es necesario identificar las utilidades de todos los pagos posibles. En la tabla 15.7 se dan, en miles de dólares, estos pagos posibles y las utilidades correspondientes. En seguida se estudiará la manera en que se obtuvieron estas utilidades.

Como punto de inicio en la construcción de la función de utilidad, puesto que tenemos la libertad de fijar el valor de $U(M)$ de forma arbitraria para dos valores de M (siempre que el valor

TABLA 15.7 Utilidades para el problema completo de la Goferbroke Co.

Pago monetario	Utilidad
−130	0
−100	0.05
60	0.30
90	0.333
670	0.97
700	1

monetario más alto tenga la mayor utilidad), fue conveniente fijar $U(-130)= 0$ y $U(700) = 1$. Después, se aplicó el *método de la lotería equivalente* para determinar la utilidad de otro de los posibles costos monetarios, $M = 90$, planteando la siguiente pregunta al tomador de decisiones (el propietario de Goferbroke Co).

> Suponga que sólo tiene las siguientes dos alternativas. En unidades de miles de dólares, la alternativa 1 significa obtener una ganancia de 700 con una probabilidad p, una ganancia de -130 (una pérdida de 130) con una probabilidad de $1 - p$. La alternativa 2, en definitiva, representa una ganancia de 90. ¿Qué valor de p haría que usted fuera *indiferente* ante estas dos alternativas?
>
> La elección del tomador de decisiones: $p = \frac{1}{3}$ por lo que $U(90) = 0.333$.

Posteriormente, el método de la lotería equivalente se aplicó de la misma forma a $M = -100$. En este caso, el *punto de indiferencia* del tomador de decisiones fue $p = \frac{1}{20}$, por lo que $U(-100) = 0.05$.

En este punto se dibujó una curva plana a través de $U(-130)$, $U(-100)$, $U(90)$ y $U(700)$ para obtener la *función de utilidad del dinero* del tomador de decisiones que se muestra en la figura 15.19. Los valores en esta curva en $M = 60$ y $M = 670$ proporcionan las utilidades correspondientes, $U(60) = 0.30$ y $U(670) = 0.97$, lo cual completa la lista de utilidades que se proporcionó en la columna derecha de la tabla 15.7. La forma de esta curva indica que el propietario de Goferbroke Co. tiene una *aversión moderada al riesgo*. En contraste, la línea punteada dibujada a 45° en la figura 15.19 muestra cómo pudo haber sido su función de utilidad si él hubiera sido *neutral ante el riesgo*.

Por su naturaleza, el dueño de la Goferbroke Co. se inclina a buscar el riesgo, pero las difíciles circunstancias financieras de su compañía, que está muy preocupado por resolver, lo han forzado a adoptar una posición de aversión moderada al riesgo para tomar la decisión actual.

Otro enfoque para estimar *U*(*M*)

El procedimiento anterior para construir $U(M)$ pide al tomador de decisiones que tome una decisión difícil varias veces acerca de la probabilidad que lo haría sentirse indiferente entre dos alternativas. Muchos individuos estarían incómodos con este tipo de decisiones. Por lo tanto, en ocasiones se usa un enfoque diferente en lugar del de la estimación de la función de utilidad del dinero.

Este enfoque se basa en suponer que la función de utilidad tiene cierta forma matemática, y después ajustar esta forma de manera que refleje la actitud del tomador de decisiones hacia el riesgo, lo más cercano posible. Por ejemplo, una forma común que se puede suponer (por su sencillez) es la **función de utilidad exponencial**,

$$U(M) = R\left(1 - e^{-\frac{M}{R}}\right),$$

donde R es la *tolerancia al riesgo* del tomador de decisiones. Esta función de utilidad muestra una utilidad marginal decreciente del dinero, por lo que está diseñada para ajustarse a la *aversión al riesgo* individual. Una gran aversión al riesgo corresponde a un valor pequeño de R (que ocasiona que la curva de la función de utilidad se doble en forma brusca), mientras que una aversión pequeña corresponde a un valor grande de R (que ocasiona un doblez gradual de la curva).

Como el dueño de Goferbroke Co. tiene una aversión al riesgo pequeña, la curva de la función de utilidad de la figura 15.19 se dobla poco a poco. Se dobla particularmente despacio para valores grandes de M cercanos al lado derecho de la figura 15.19, por lo que el valor correspondiente de R en esta región es aproximadamente $R = 2\ 000$. Por otro lado, el propietario muestra mucha más aversión al riesgo cuando se pueden presentar grandes pérdidas, ya que esto lo podría dejar en la bancarrota, por lo que la curva de la función de utilidad tiene una curvatura considerablemente mayor en esta región donde M tiene valores negativos grandes. Por lo tanto, el valor correspondiente de R es considerablemente menor sólo alrededor de $R = 500$, en esta región.

Desafortunadamente, no es posible usar dos valores diferentes de R para la misma función de utilidad. Una desventaja de la función de utilidad exponencial es que supone una aversión al riesgo constante (valor fijo de R), sin que importe cuánto dinero tiene en la actualidad el tomador de decisiones. Esto no se ajusta a la situación de Goferbroke porque la falta de dinero actual obliga al dueño a preocuparse más que lo usual por una pérdida grande.

En situaciones donde las consecuencias de pérdidas potenciales no son tan severas, el supuesto de una función de utilidad exponencial puede proporcionar una aproximación razonable. En ese caso, existe una manera sencilla (aproximada) de estimar el valor apropiado de R. Se pide al tomador de decisiones que elija el número R que le haría sentir indiferente entre las dos alternativas siguientes.

A_1: Un juego de 50-50 donde ganaría R dólares con probabilidad 0.5 y perdería $\frac{R}{2}$ dólares con probabilidad 0.5.

A_2: No perdería ni ganaría nada.

TreePlan incluye la opción de usar la función de utilidad exponencial. Primero, el valor de R debe especificarse en la hoja de cálculo. La celda que contiene este valor necesita asignársele un nombre de rango de RT (TreePlan se refiere a este término como la tolerancia al riesgo). Después haga click en el botón Options en el cuadro de diálogo TreePlan y seleccione la opción “Use Exponential Utility Function”. Al presionar OK se efectúa una revisión del árbol de decisiones para incorporar la función de utilidad exponencial.

Uso de un árbol de decisión para analizar el problema de Goferbroke Co. con utilidades

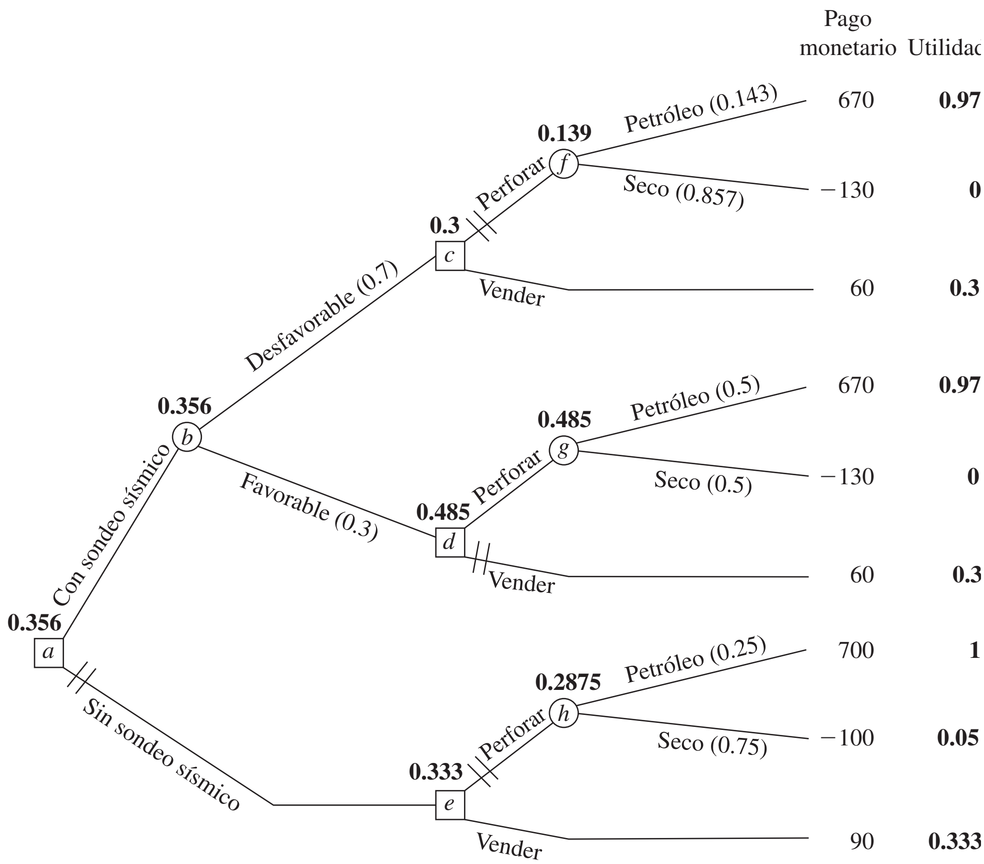
Una vez que se obtiene la función de utilidad del dinero que corresponde al dueño de Goferbroke Co. en la tabla 15.7 (y la figura 15.19), esta información se puede usar en un árbol de decisión como se resume a continuación.

> El procedimiento para usar un árbol de decisión para analizar el problema es *idéntico* al descrito en la sección anterior *excepto* que se sustituyen los pagos monetarios por las utilidades. Por lo tanto, el valor que se obtiene de evaluar cada nodo del árbol es ahora la *utilidad esperada* en lugar del pago esperado (monetario). En consecuencia, las decisiones óptimas que seleccionan la regla de decisión de Bayes maximizan la utilidad esperada para todo el problema.

Así, el árbol de decisión final que se muestra en la figura 15.20 se parece al de la figura 15.6 dado en la sección 15.4. Los nodos y las ramas son los mismos, al igual que las probabilidades de las ramas que salen de los nodos de probabilidad. Con fines informativos, los pagos monetarios totales todavía están dados a la derecha de las ramas terminales (pero ya no muestran los pagos monetarios individuales al lado de las ramas). Sin embargo, ahora se han agregado las utilidades en el lado derecho. Son estos números los que se usaron para calcular las utilidades esperadas que se observan cerca de cada nodo.

Estas utilidades esperadas llevan a las mismas decisiones de los nodos *a*, *c* y *d* que en la figura 15.6, pero la decisión del nodo *e* cambia ahora a *vender* en lugar de *perforar*. Sin embargo, el procedimiento de inducción hacia atrás todavía deja al nodo *e* en una trayectoria *cerrada*. Por lo tanto, la política óptima global es la misma que la dada al final de la sección 15.4 (hacer el sondeo sísmico; vender si el resultado es desfavorable; perforar si el resultado es favorable).

El enfoque que se utilizó en las secciones anteriores es maximizar las cantidades del pago monetario esperado si es que el tomador de decisiones es neutral al riesgo, de manera que $U(M) = M$. Con base en la teoría de la utilidad, la solución óptima ahora refleja la actitud del tomador



■ **FIGURA 15.20**
Árbol de decisiones final para el problema de la Goferbroke Co., usando la función de utilidad del dinero del dueño con el fin de maximizar la utilidad esperada.

de decisiones respecto al riesgo. Como el dueño de la Goferbroke Co. adoptó sólo una posición de aversión moderada al riesgo, la política óptima no cambia respecto de la anterior. En el caso de una persona con mayor aversión al riesgo, la solución óptima cambiaría a la más conservadora de vender el terreno de inmediato (sin exploración). (Vea el problema 15.6-1.)

El dueño actual merece un reconocimiento por incorporar la teoría de la utilidad al enfoque de análisis de decisiones de su problema, pues ella ayuda a elaborar un enfoque racional de la toma de decisiones frente a la incertidumbre. Sin embargo, muchos tomadores de decisiones no se sienten a gusto con la noción un tanto abstracta de la utilidad o con las probabilidades para construir la función de utilidad como para querer usarlo. En consecuencia, la teoría de la utilidad todavía no tiene una aplicación muy amplia en la práctica.

15.7 APLICACIÓN PRÁCTICA DEL ANÁLISIS DE DECISIONES

En un sentido, el ejemplo prototipo de este capítulo (problema de Goferbroke Co.) es una aplicación común del análisis de decisión. Como otras aplicaciones, la administración necesita tomar algunas decisiones (realizar el estudio, perforar o vender) ante una gran incertidumbre. Las decisiones son difíciles porque sus pagos son impredecibles. El resultado depende de factores que están fuera del control del administrador (¿el terreno contiene petróleo o está seco?). Por lo tanto, la administración necesita un marco de trabajo y una metodología para tomar decisiones racionales en este entorno de incertidumbre. Éstas son las características usuales de las aplicaciones del análisis de decisión.

Sin embargo, en otro sentido, el problema de Goferbroke no es una aplicación tan usual. Se simplificó mucho para que incluyera sólo dos estados posibles de la naturaleza (petróleo y

seco), mientras que en realidad habría un gran número de posibilidades. Por ejemplo, el estado real puede ser seco, poco petróleo, cantidad moderada, mucho petróleo y una gran cantidad de petróleo, además de varias posibilidades referentes a la profundidad a la que se encuentra y las condiciones del suelo que afectan el costo de la perforación para llegar al hidrocarburo. La administración también considera dos alternativas para cada una de las dos decisiones. Las aplicaciones reales suelen involucrar más decisiones, más alternativas para cada una y muchos estados posibles de la naturaleza.

Cuando se manejan problemas más grandes, el árbol de decisión puede explotar en tamaño, tal vez con muchos miles de ramas terminales. En este caso, es claro que no sería factible construir el árbol a mano, calcular las probabilidades *a posteriori* y de los pagos esperados (o utilidades) de todos los nodos, y después identificar las decisiones óptimas. Por fortuna, se dispone de excelentes paquetes de software (casi todos para computadoras personales), especiales para hacer este trabajo. Aún más, se están desarrollando técnicas algebraicas especiales que se incorporan a los solucionadores de computadora para manejar problemas cada vez más grandes.[4]

El análisis de sensibilidad también puede ser difícil de manejar en problemas grandes. Aunque casi siempre tiene apoyo de la computadora, la cantidad de datos generados puede fácilmente abrumar al analista o tomador de decisiones. Se han desarrollado algunas técnicas gráficas, como los *diagramas tornado*, para organizar los datos en una forma fácil de entender.[5]

También se dispone de otras técnicas gráficas para complementar el árbol de decisión que representan y resuelven los problemas de análisis de decisiones. Una bastante conocida se llama *diagrama de influencia* y los investigadores continúan con el desarrollo de otras.[6]

Muchas decisiones estratégicas de negocios se toman de manera colectiva entre varios miembros de la administración. Una técnica para hacerlo en grupo se llama *conferencia de decisiones*. En este proceso, un grupo se reúne para analizar una decisión en conferencia con la ayuda de un analista y un facilitador. Éste trabaja en contacto directo con el grupo para ayudar a estructurar y centrar las discusiones, pensar con creatividad en el problema, elaborar los supuestos y estudiar la gama completa de aspectos involucrados. El analista usa el análisis de decisiones para ayudar al grupo a explorar las implicaciones de las decisiones alternativas. Con la asistencia de un sistema de apoyo computarizado para tomar decisiones en grupo, el analista construye y resuelve los modelos ahí mismo, y después realiza un análisis de sensibilidad para responder a las preguntas “qué pasa si” del grupo.[7]

Las aplicaciones del análisis de decisiones, por lo común, incluyen la participación de los tomadores de decisiones de la administración (ya sea individual o en grupo) y un analista (individual o un equipo) con conocimientos de IO. Algunas compañías no cuentan con personal calificado para actuar como analista. Por esto se ha formado un número considerable de empresas consultoras que se especializan en análisis de decisiones para dar este servicio.

Si desea leer más sobre las aplicaciones prácticas del análisis de decisiones, un buen sitio para comenzar es el número de noviembre-diciembre de 1992 de *Interfaces*. Es un número especial dedicado por completo al análisis de decisiones y al área relacionada de análisis de riesgo. Incluye muchos artículos interesantes, con descripciones de los métodos básicos, análisis de sensibilidad y conferencias de decisión. También contiene varios artículos sobre las aplicaciones. Entonces, para obtener una perspectiva más reciente acerca de la aplicación práctica del análisis de decisiones, sugerimos que consulte la referencia seleccionada 8. Este artículo fue el primero en el primer número de la publicación *Decision Analysis* que se enfocaba en la investigación aplicada en el análisis de decisiones. El artículo proporciona un análisis detallado de varias publicaciones que presentan aplicaciones de este tema.

[4] Por ejemplo, vea C. W. Kirkwood, “An Algebraic Approach to Formulating and Solving Large Models for Sequential Decisions under Uncertainty”, en *Management Science*, **39:** 900-913, julio de 1993.

[5] Para mayor información, vea T. G. Eschenbach, “Spiderplots versus Tornado Diagrams for Sensitivity Analysis”, en *Interfaces*, **22:** 40-46, noviembre-diciembre de 1992.

[6] Por ejemplo, consulte C. Bielza y P. P. Schnoy, “A Comparison of Graphical Techniques for Asymmetric Decision Problems”, en *Management Science*, **45**(11): 1552-1569, noviembre de 1999.

[7] Para mayor información, vea los dos artículos sobre conferencias de decisión en el número correspondiente a noviembre-diciembre 1992 de *Interfaces*, donde uno describe una aplicación en Australia y el otro resume las experiencias de 26 conferencias de decisión en Hungría.

15.8 CONCLUSIONES

El análisis de decisión se ha convertido en una técnica importante para la toma de decisiones bajo incertidumbre. Se caracteriza por la enumeración de todos los cursos de acción disponibles, identifica los pagos de todos los resultados posibles y cuantifica las probabilidades subjetivas de todos los eventos aleatorios. Cuando se cuenta con estos datos, el análisis de decisión se convierte en una herramienta poderosa para determinar un curso de acción óptimo.

Una opción que se puede incorporar con facilidad al análisis es llevar a cabo una experimentación para obtener mejores estimaciones de las probabilidades de todos los estados posibles de la naturaleza. Los árboles de decisión son una herramienta visual útil para analizar esta opción o cualquier serie de decisiones.

La teoría de la utilidad proporciona una manera de incorporar al análisis la actitud del tomador de decisiones frente al riesgo.

Cada vez se dispone de más paquetes de computadora para realizar el análisis de decisiones (se incluyen TreePlan y SensIt en el OR Courseware). (La referencia seleccionada 9 proporciona un estudio de dicho software.)

REFERENCIAS SELECCIONADAS

1. Bleichrodt, H., J. M. Abellan-Perpiñan, J. L. Pinto-Prades e I. Mendez-Martinez: "Resolving Inconsistencies in Utility Measurement Under Risk: Test of generalizations of Expected Utility", en *Management Science*, **53**(3): 469-482, marzo 2007.
2. Clemen, R. T.: *Making Hard Decisions: Introduction to Decision Analysis* (with CD-ROM), 3a. ed., Duxbury Press, Pacific Grove, CA, 2006.
3. Fishburn, P. C.: "Foundations of Decision Analysis: Along the Way", *Management Science*, **35:** 387-405, 1989.
4. Fishburn, P. C.: *Nonlinear Preference and Utility Theory*, The Johns Hopkins Press, Baltimore, MD, 1988.
5. Goodwin, P. y G. Wright: *Decision Analysis for Management Judgment*, Wiley, Nueva York, 2004.
6. Hammond, J. S., R. L. Keeney y H. Raiffa: *Smart Choices: A Practical Guide to Making Better Decisions*, Harvard Business School Press, Cambridge, MA, 1999.
7. Hillier, F. S. y M. S. Hillier: *Introduction to Management Science: A Modeling and Case Studies Approach with Spreadsheets*, 3a. ed., McGraw-Hill/Irwin, Burr Ridge, IL, 2008, cap. 9.
8. Keefer, D. L., C. W. Kirkwood y J. L. Corner: "Perspective on Decision Analysis Applications", en *Decision Analysis*, **1**(1): 4-22, 2004.
9. Maxwell, D. T.: "Software Survey: Decision Analysis", en *OR/MS Today*, **33**(6):51-61, diciembre de 2006.
10. Smith, J. E. y R. L. Keeney: "Your Money or Your Life: A Prescriptive Model for Health, Seafty and Consumption Decisions", en *Management Science*, **51**(9): 1309-1325, septiembre de 2005.
11. Smith, J. E. y R. L. Winkler: "The Optimizer´s Curse: Skepticism and Postdecision Surprise in Decision Analysis", en *Management Science*, **52**(3): 311-322, marzo de 2006.
12. Smith, J. E. y D. von Winterfeldt: "Decision Analysis in Management Science", en *Management Science*, **50**(5): 561-574, mayo de 2004.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos del capítulo 15

Archivos de Excel "Ch 15—Decision Analysis":

Plantilla para probabilidades *a posteriori*
Árbol de decisión TreePlan para el primer problema de Goferbroke Co.
Árbol de decisión TreePlan para el problema completo de Goferbroke (con gráficos de SensIt)

Archivo de LINGO para los ejemplos seleccionados "Ch 15—Decision Analysis":

Complementos de Excel:

TreePlan (versión académica)
SensIt (versión académica)

Glosario para el capítulo 15

Vea en el apéndice 1 la documentación del software.

PROBLEMAS

Los símbolos a la izquierda de algunos problemas (o de sus incisos) significan lo siguiente:

T: La plantilla de Excel correspondiente puede ser útil.
A: El complemento de Excel correspondiente mencionado con anterioridad puede ser útil.

Un asterisco en el número del problema indica que al final del libro se da al menos una respuesta parcial.

15.2-1. Lea el artículo de referencia que describe el estudio de IO que se resume en el recuadro de aplicación que se presentó en la sección 15.2. Describa de manera breve la forma en que el análisis de decisiones se aplicó a este estudio. Después, elabore una lista de los beneficios financieros y no financieros que arrojó dicho estudio.

15.2-2.* Silicon Dynamics diseñó un nuevo circuito integrado que le permitirá entrar, si así lo desea, al campo de las microcomputadoras. De otra manera, puede vender sus derechos por 15 millones de dólares. Si elige construir computadoras, la rentabilidad de este proyecto depende de la habilidad de la compañía para comercializarlas durante el primer año. Tiene suficiente acceso a los distribuidores al menudeo como para garantizar la venta de 10 000 de ellas. Por otro lado, si tiene éxito puede llegar a vender hasta 100 000 unidades. Con propósitos de análisis, estos dos niveles de ventas se toman como dos resultados posibles de la venta de computadoras. El costo de instalar la línea de producción es de 6 millones de dólares. La diferencia entre el precio de venta y el costo variable de cada computadora es de 600 dólares.

***a*)** Desarrolle una formulación de análisis de decisiones para este problema mediante la identificación de las acciones, los estados de la naturaleza y la matriz de pagos.
***b*)** Desarrolle una gráfica del pago esperado para cada acción alternativa contra la probabilidad *a priori* de vender 10 000 computadoras.
***c*)** Respecto de la gráfica que desarrolló en el inciso *b*), use el álgebra para obtener el *punto de cruce*. Explique el significado de este punto.
A ***d*)** Desarrolle una gráfica del pago esperado (con la regla de decisión de Bayes) contra la probabilidad *a priori* de vender 10 000 computadoras.
***e*)** Suponga que ambas probabilidades *a priori* de los dos niveles de ventas son iguales a 0.5. ¿Qué alternativa de acción debe elegirse?

15.2-3. Jean Clark es la gerente de Midtown Saveway Grocery Store, empresa que necesita reabastecer su inventario de fresas. Su proveedor normal puede surtir todas las cajas que desee. Sin embargo, como ya están muy maduras, deberá venderlas el día siguiente y después desechar las que queden. Jean estima que podrá vender 12, 13, 14 o 15 cajas mañana. Puede comprar las fresas en 7 dólares por caja y venderlas en 18 dólares. Jean ahora necesita decidir cuántas cajas comprará.

Jean verifica los registros de ventas diarias de fresas de la tienda. Con base en ellos, estima que las probabilidades *a priori* de poder vender 12, 13, 14 y 15 cajas de fresas mañana son 0.1, 0.3, 0.4 y 0.2, respectivamente.

***a*)** Desarrolle la formulación del análisis de decisión de este problema mediante la identificación de las acciones alternativas, los estados de la naturaleza y la tabla de pagos.
***b*)** ¿Cuántas cajas de fresas debe comprar Jean si se basa en el criterio de pago máximo?
***c*)** ¿Cuántas cajas debe comprar según el criterio de la máxima posibilidad?
***d*)** ¿Cuántas cajas debe comprar según la regla de decisión de Bayes?
***e*)** Jean piensa que las probabilidades *a priori* para la venta de 12 y 15 cajas son correctas, pero no está segura de cómo dividir esas probabilidades para 13 y 14 cajas. Aplique de nuevo la regla de decisión de Bayes cuando las probabilidades *a priori* de vender 13 y 14 cajas son: *i*) 0.2 y 0.5, *ii*) 0.4 y 0.3 y *iii*) 0.5 y 0.2.

15.2-4.* Warren Buffy es un inversionista muy rico que ha amasado su fortuna con su legendaria perspicacia y quiere hacer una inversión. La primera opción es una *inversión conservadora* con buen desempeño si la economía mejora y sólo sufrirá una pérdida pequeña si la economía empeora. La segunda es una *inversión especulativa* que se desempeña muy bien si la economía mejora, pero muy mal si empeora. La tercera es una *inversión contracíclica* que perdería algún dinero en una economía que mejora, pero se desempeñaría muy bien si empeora.

Warren cree que existen tres escenarios posibles en las vidas de estas inversiones potenciales: 1) economía que mejora, 2) economía estable y 3) economía que empeora. Es pesimista sobre el rumbo de la economía, y ha asignado probabilidades *a priori* respectivas de 0.1, 0.5 y 0.4, a estos tres escenarios. También estima que sus ganancias en estos escenarios son las que se presentan en la tabla siguiente:

	Economía en mejoría	Economía estable	Economía que empeora
Inversión conservadora	\$30 millones	\$ 5 millones	−\$10 millones
Inversión especulativa	\$40 millones	\$10 millones	−\$30 millones
Inversión contracíclica	−\$10 millones	0	\$15 millones
Probabilidad *a priori*	0.1	0.5	0.4

¿Qué inversión debe hacer Warren según los siguientes criterios?

a) Criterio de pago maximin.
b) Criterio de la posibilidad máxima.
c) Regla de decisión de Bayes.

15.2-5. Reconsidere el problema 15.2-4. Warren Buffy decide que la regla de decisión de Bayes es el criterio más confiable. Cree que 0.1 es correcto como probabilidad *a priori* de una economía que mejora, pero no sabe cómo dividir el resto de las probabilidades entre la economía estable y la que empeora. Por lo tanto, quiere realizar un análisis de sensibilidad respecto de estas dos probabilidades.

a) Aplique de nuevo la regla de decisión de Bayes cuando la probabilidad *a priori* de una economía estable es 0.3 y la de una economía que empeora es 0.6.
b) Aplique de nuevo la regla de decisión de Bayes cuando la probabilidad *a priori* de una economía estable es 0.7 y la de una economía que empeora es 0.2.
c) Grafique la ganancia esperada de las tres alternativas de inversión contra la probabilidad *a priori* de una economía estable (con probabilidad *a priori* de economía que mejora fija en 0.1). Use la gráfica para identificar el punto de cruce donde la decisión cambia de una inversión a otra.
d) Use el álgebra para obtener los puntos de cruce que se mencionan en el inciso *c*).
A ***e***) Desarrolle una gráfica de la ganancia esperada (con la regla de decisión de Bayes) contra la probabilidad *a priori* de una economía estable.

15.2-6. Considere un problema de análisis de decisión cuyos pagos (en miles de dólares) se dan en la siguiente tabla de pagos:

Alternativa	Estado de la naturaleza		
	S_1	S_2	S_3
A_1	220	170	110
A_2	200	180	150
Probabilidad *a priori*	0.6	0.3	0.1

a) ¿Qué alternativa debe elegir según el criterio de pago maximin?
b) ¿Qué alternativa debe elegir según el criterio de la máxima posibilidad?
c) ¿Qué alternativa debe elegir según la regla de decisión de Bayes?
d) Use la regla de decisión de Bayes para hacer un análisis de sensibilidad gráfico respecto de las probabilidades *a priori* de los estados S_1 y S_2 (sin cambiar la probabilidad *a priori* del estado S_3) y determinar los puntos de cruce donde la decisión cambia de una alternativa a otra. Después use el álgebra para obtener estos puntos de cruce.
e) Repita el inciso *d*) para las probabilidades *a priori* de los estados S_1 y S_3.
f) Repita el inciso *d*) para las probabilidades *a priori* de los estados S_2 y S_3.
g) Si usted siente que las probabilidades verdaderas de los estados de la naturaleza están dentro de 10% de las probabilidades *a priori* dadas, ¿qué alternativa debería elegir?

15.2-7. Dwight Moody es el administrador de un rancho con 1 000 hectáreas de tierra cultivable. Para mayor eficiencia, Dwight siempre cosecha un tipo de cultivo a la vez. Ahora debe decidir entre cuatro cultivos para la próxima temporada. En el caso de cada cultivo ha obtenido las siguientes estimaciones sobre la cosecha y los precios por bushel con diferentes condiciones del clima:

Clima	Utilidad esperada, bushels/acre			
	Cultivo 1	Cultivo 2	Cultivo 3	Cultivo 4
Seco	30	25	40	60
Moderado	50	30	35	60
Húmedo	60	40	35	60
Ingreso neto por bushel	\$3.00	\$4.50	\$3.00	\$1.50

Después de estudiar los registros meteorológicos, estima las siguientes probabilidades *a priori* sobre el clima durante la temporada:

Seco	0.2
Moderado	0.5
Húmedo	0.3

a) Desarrolle una formulación para el análisis de decisiones de este problema mediante la identificación de las acciones, los estados de la naturaleza y la matriz de pagos.
b) Utilice la regla de decisión de Bayes para determinar cuál cosecha plantar.
c) Use la regla de decisión de Bayes para realizar un análisis de sensibilidad respecto de las probabilidades *a priori* de clima moderado y clima húmedo (sin cambiar la probabilidad *a priori* de clima seco) al resolver de nuevo cuando la probabilidad *a priori* de clima moderado es 0.2, 0.3, 0.4 y 0.6.

15.2-8.* La Fuerza Aérea comprará un nuevo tipo de avión y debe determinar el número de motores de repuesto que va a ordenar. Estos motores de repuesto se deben ordenar en lotes de cinco y se puede elegir sólo entre 15, 20 o 25 repuestos. El proveedor de motores tiene dos plantas y la Fuerza Aérea debe tomar sus decisiones antes de saber qué planta se usará. Por experiencia, la Fuerza Aérea sabe que dos tercios de los motores de todos tipos se producen en la planta *A* y sólo un tercio en la planta *B*. La Fuerza Aérea también sabe que el número de motores de repuesto requeridos cuando se lleva a cabo la producción en la planta *A* se aproxima por una distribución Poisson con media $\theta = 21$, mientras que el número de motores de repuesto que se requiere cuando la producción se lleva a cabo en la planta *B* se

aproxima por una distribución Poisson con media $\theta = 24$. El costo de un motor de repuesto comprado ahora es de 400 000 dólares, pero si se compra después será de 900 000 dólares. Los repuestos siempre se surten si se piden y las máquinas que no se usen serán chatarra cuando los aviones caigan en la obsolescencia. Los costos de mantener un inventario y del interés pueden despreciarse. A partir de estos datos se calculó el costo total (pagos negativos) de la siguiente forma:

	Estado de la naturaleza	
Alternativa	$\theta = 21$	$\theta = 24$
Ordenar 15	1.155×10^7	1.414×10^7
Ordenar 20	1.012×10^7	1.207×10^7
Ordenar 25	1.047×10^7	1.135×10^7

Determine la acción óptima según la regla de decisión de Bayes.

15.3-1. Lea el artículo de referencia que describe el estudio de IO que se resume en el recuadro de aplicación que se presentó en la sección 15.3. Describa de manera breve la forma en que el análisis de decisiones se aplicó a este estudio. Después, elabore una lista de los beneficios financieros y no financieros que arrojó dicho estudio.

15.3-2.* Reconsidere el problema 15.2-2. La administración de Silicon Dynamics considera ahora realizar un estudio de mercado a un costo de 1 millón de dólares para predecir cuál de los dos niveles de demanda es más probable que ocurra. La experiencia indica que esta investigación de mercado es correcta dos tercios del tiempo.

a) Encuentre el VEIP de este problema.
b) ¿La respuesta en el inciso *a*) indica que vale la pena realizar la investigación de mercado?
c) Desarrolle el diagrama de árbol de probabilidad para obtener las probabilidades *a posteriori* de los dos niveles de demanda para los dos resultados posibles de la investigación de mercado.
T ***d)*** Use la plantilla de Excel correspondiente para verificar sus respuestas en *c*).
e) Encuentre el VEE. ¿Vale la pena realizar la investigación de mercado?

15.3-3. Se proporciona la siguiente matriz de pagos (en miles de dólares) para un análisis de decisión:

	Estado de la naturaleza		
Alternativa	S_1	S_2	S_3
A_1	6	1	1
A_2	1	3	0
A_3	4	1	2
Probabilidad *a priori*	0.3	0.4	0.3

a) De acuerdo con la regla de decisión de Bayes, ¿qué alternativa debe elegirse?
b) Encuentre el VEIP.
c) Se le proporciona la oportunidad de gastar 1 000 dólares para obtener información acerca de qué estado de la naturaleza es probable que ocurra. Dada su respuesta del inciso *b*), ¿vale la pena gastar este dinero?

15.3-4.* Betsy Pitzer toma decisiones según la regla de decisión de Bayes. Para su problema actual, Betsy construye la siguiente tabla de pagos (en dólares):

	Estado de la naturaleza		
Alternativa	S_1	S_2	S_3
A_1	50	100	−100
A_2	0	10	−10
A_3	20	40	−40
Probabilidad *a priori*	0.5	0.3	0.2

a) ¿Qué alternativa debe elegir Betsy?
b) Encuentre el VEIP.
c) ¿Cuál es el gasto máximo que Betsy debe considerar para obtener más información acerca de qué estado de la naturaleza ocurrirá?

15.3-5. Por medio de la regla de decisión de Bayes, considere el problema de análisis de decisión con la siguiente matriz de pagos (en miles de dólares):

	Estado de la naturaleza		
Alternativa	S_1	S_2	S_3
A_1	−20	3	25
A_2	−3	5	10
A_3	4	2	15
Probabilidad *a priori*	0.3	0.3	0.4

a) ¿Qué alternativa debe elegirse? ¿Cuál es el pago esperado que resulta?
b) Se tiene la oportunidad de obtener información que dirá con certidumbre si ocurrirá el primer estado de la naturaleza S_1. ¿Cuál es la cantidad máxima que debe pagarse por la información? Suponga que obtiene la información. ¿Cómo debe usarla para elegir una alternativa?
c) Ahora repita el inciso *b*) si la información ofrecida se refiere a S_2 en lugar de S_1.
d) Ahora repita el inciso *b*) si la información ofrecida se refiere a S_3 en lugar de S_1.
e) Ahora suponga que se ofrece la oportunidad de proporcionar información que dirá con certidumbre qué estado de la naturaleza ocurrirá (información perfecta). ¿Cuál es la cantidad máxima que debe pagarse por la información? Suponga que obtiene la información. ¿Cómo debe usarla para elegir una alternativa? ¿Cuál es el pago esperado (sin incluir la información)?
f) Si tiene la oportunidad de hacer algunas pruebas que le den información parcial adicional (no la información perfecta) sobre el estado de la naturaleza, ¿cuál es la cantidad máxima que debe pagar por ella?

15.3-6. Reconsidere el ejemplo prototipo de Goferbroke Co. que incluye el análisis de la sección 15.3. Con la ayuda de un geólogo consultor se han obtenido algunos datos históricos que proporcionan más información precisa de la posibilidad de obtener sondeos favorables en terrenos similares. En específico, cuando la tierra contiene

petróleo, los sondeos favorables se confirman 80% de las veces. Este porcentaje cambia a 40 cuando el terreno está seco.

a) Revise la figura 15.2 para encontrar las nuevas probabilidades *a posteriori*.

T ***b***) Use la plantilla correspondiente de Excel para verificar sus respuestas en el inciso *a*).

c) ¿Cuál es la política óptima que resulta?

15.3-7. Se da la siguiente tabla de pagos (en dólares):

Alternativa	Estado de la naturaleza S_1	S_2
A_1	400	−100
A_2	0	100
Probabilidad *a priori*	0.4	0.6

Se tiene la opción de pagar 100 dólares para hacer una investigación para predecir mejor qué estado de la naturaleza ocurrirá. Cuando el estado de la naturaleza es S_1, la investigación predice con exactitud S_1 60% del tiempo (pero erróneamente predice S_2 40% del tiempo). Cuando el estado de la naturaleza es S_2, la investigación predice bien S_2 80% del tiempo (pero predice S_1 20% del tiempo).

a) Dado que no se hace la investigación, use la regla de decisión de Bayes para determinar qué alternativa debe elegirse.

b) Encuentre el VEIP. ¿Indica la respuesta que quizá valga la pena hacer la investigación?

c) Dado que se realiza la investigación, encuentre la probabilidad conjunta de los siguientes pares de resultados: *i*) el estado de la naturaleza es S_1 y la investigación predice S_1, *ii*) el estado es S_1 y se predice S_2, *iii*) el estado de la naturaleza es S_2 y se predice S_1 y *iv*) el estado de la naturaleza es S_2 y la investigación predice S_2.

d) Encuentre la probabilidad no condicional de que la investigación prediga S_1. También encuentre la probabilidad no condicional de que prediga S_2.

e) Dado que se realiza el estudio, use sus respuestas en *c*) y *d*) para determinar las probabilidades *a posteriori* de los estados de la naturaleza para las dos predicciones posibles de la investigación.

T ***f***) Use la plantilla de Excel correspondiente para obtener las respuestas del inciso *e*).

g) Dado que el estudio predice S_1, use la regla de decisión de Bayes para determinar qué alternativa de decisión debe elegirse y el pago esperado.

h) Repita el inciso *g*) cuando el estudio predice S_2.

i) Dado que se realiza el estudio, ¿cuál es el pago esperado cuando se usa la regla de decisión de Bayes?

j) Use los resultados anteriores para determinar la política óptima respecto a si realiza la investigación y la elección de la alternativa de decisión.

15.3-8.* Reconsidere el problema 15.2-8. Suponga que la Fuerza Aérea sabe que se produjo un tipo parecido de máquina para una versión anterior del avión que ahora se considera. El tamaño de la orden para el modelo anterior fue el mismo que para el actual. Además, se piensa que la distribución de probabilidad del número de motores de repuesto necesarios, dada la planta en donde se fabricó, es la misma para el modelo anterior y para el actual. El motor para surtir la orden actual se producirá en la misma planta que el modelo anterior, aunque la Fuerza Aérea no sabe cuál es. Ellos tienen acceso a los datos del número de motores de repuesto que se requieren para la versión anterior, pero el proveedor no ha revelado dónde los fabricará.

a) ¿Cuánto dinero vale la pena pagar por la información perfecta sobre la planta que producirá los motores?

b) Suponga que el costo de los datos sobre el modelo anterior de avión es cero y que se requirieron 30 motores de repuesto. Se sabe que la probabilidad de 30 repuestos, dada una distribución Poisson con media θ, es 0.013 para $\theta = 21$ y 0.036 para $\theta = 24$. Encuentre la opción óptima según la regla de decisión de Bayes.

15.3-9.* Vincent Cuomo es el gerente de crédito de Fine Fabrics Mill y se enfrenta al problema de extender un crédito de 100 000 dólares a uno de sus nuevos clientes, un fabricante de vestidos. Vincent clasifica a sus clientes en tres categorías: riesgo malo, riesgo promedio y riesgo bueno, pero no sabe en qué categoría puede colocar a este nuevo cliente. Su experiencia indica que 20% de las compañías semejantes se consideran riesgo malo, 50% son riesgo promedio y 30% son riesgo bueno. Si se extiende el crédito, la ganancia esperada para las de riesgo malo llega a –15 000 dólares, para las de riesgo promedio es de 10 000 dólares y para las de riesgo bueno es de 20 000 dólares. Si no extiende el crédito, el fabricante de vestidos recurrirá a otro fabricante textil. La fábrica puede consultar a una organización dedicada a la clasificación de créditos que cobra 5 000 dólares por compañía evaluada. Para las compañías con historial crediticio vigente incluido en cada una de las tres categorías, la siguiente tabla muestra los porcentajes dada a cada una de las tres posibles evaluaciones de crédito de la organización:

Evaluación de créditos	Clasificación real del crédito: Malo	Promedio	Bueno
Malo	50%	40%	20%
Promedio	40%	50%	40%
Bueno	10%	10%	40%

a) Desarrolle una formulación para el análisis de decisiones de este problema mediante la identificación de las opciones posibles y los estados de la naturaleza y después construya la matriz de pagos.

b) ¿Cuál es la opción óptima según la regla de decisión de Bayes si se supone que no se consulta a la organización de evaluación de créditos?

c) Encuentre el VEIP. ¿Indica esta respuesta que se debe recurrir a la organización de evaluación?

d) Suponga que se recurre a la organización de evaluación de créditos. Desarrolle un diagrama de árbol para encontrar las probabilidades *a posteriori* de los respectivos estados de la naturaleza para las tres evaluaciones de crédito posibles de este cliente potencial.

T ***e***) Use la plantilla de Excel correspondiente para obtener las respuestas del inciso *d*).

f) Determine la política óptima de Vincent.

15.3-10. Una liga deportiva realiza pruebas antidoping de sus atletas, 15% de los cuales usan drogas. Esta prueba tiene sólo 97% de confiabilidad. Es decir, un atleta que usa drogas saldrá positivo con probabilidad 0.97 y negativo con probabilidad 0.03, y aquel que no consume estupefacientes saldrá negativo y positivo con probabilidades respectivas de 0.97 y 0.03.

Desarrolle un árbol para determinar la probabilidad *a posteriori* de los resultados de someter a prueba a un atleta.

***a*)** Es usuario de drogas, dado que la prueba es positiva.
***b*)** No es usuario, dado que la prueba es positiva.
***c*)** Es usuario de drogas dado que la prueba es negativa.
***d*)** No es usuario, dado que la prueba es negativa.
T ***e*)** Use la plantilla de Excel correspondiente para verificar sus respuestas a los incisos anteriores.

15.3-11. La administración de la compañía Telemore estudia el desarrollo y comercialización de un nuevo producto. Se estima que hay el doble de posibilidades de que el producto tenga éxito a que no lo tenga. Si tiene éxito, la ganancia esperada sería $1 500 000. Si no es así, la pérdida esperada sería $1 800 000. Se puede hacer una investigación de mercado a un costo de 300 000 dólares para predecir si tendrá o no éxito. La experiencia indica que se ha pronosticado éxito de productos exitosos 80% del tiempo y fracaso de productos no exitosos 70% del tiempo.

***a*)** Desarrolle una formulación para el análisis de decisión de este problema mediante la identificación de las opciones alternativas, los estados de la naturaleza y la matriz de pagos cuando se realiza el estudio de mercado.
***b*)** Suponga que no se realiza el estudio de mercado; use la regla de decisión de Bayes para determinar qué alternativa debe elegirse.
***c*)** Encuentre el VEIP. ¿Indica esta respuesta que debe tomarse en cuenta la realización del estudio de mercado?
T ***d*)** Suponga que se realiza el estudio de mercado. Encuentre las probabilidades *a posteriori* de los respectivos estados de las dos predicciones posibles del estudio de mercado.
***e*)** Encuentre la política óptima para determinar si se debe realizar el estudio y si se debe desarrollar y vender el nuevo producto.

15.3-12. La compañía Hit-and-Miss produce artículos que tienen una probabilidad p de salir defectuosos. Se forman lotes de 150 artículos con ellos. La experiencia indica que el valor de p es 0.05 o 0.25 y que en 80% de los lotes que se producen p es igual a 0.05 (de manera que p es igual a 0.25 en 20% de los lotes). Estos artículos se utilizan después en un ensamble y, en última instancia, su calidad se determina antes de que el producto final salga de la planta. En principio el fabricante puede ya sea *inspeccionar* cada artículo del lote con un costo de 10 dólares por artículo y reemplazar los defectuosos, o bien utilizarlos sin inspección. Si se elige esta acción, el costo al tener que volver a hacer el ensamble es de 100 dólares por artículo defectuoso. Como la inspección debe programar inspectores y equipo, la decisión de realizarla o no debe tomarse dos días antes. Sin embargo, se puede tomar un artículo de un lote e inspeccionarlo; su calidad (defectuoso o aceptable) se informa antes de tomar la decisión de inspeccionar o no. El costo de esta inspección inicial es de 125 dólares.

***a*)** Desarrolle una formulación para el análisis de decisión de este problema, e identifique las acciones alternativas, los estados de la naturaleza y la matriz de pagos si no se inspecciona un artículo de antemano.
***b*)** Suponga que no se inspecciona un artículo de antemano; use la regla de decisión de Bayes para determinar qué alternativa debe elegirse.
***c*)** Encuentre el VEIP. ¿Indica esta respuesta que debe considerarse inspeccionar el artículo de antemano?
T ***d*)** Suponga que se inspecciona el artículo de antemano. Encuentre las probabilidades *a posteriori* de los respectivos estados de la naturaleza de los dos resultados posibles de esta inspección.
***e*)** Encuentre VEE. ¿Vale la pena inspeccionar el artículo?
***f*)** Determine la política óptima.

T **15.3-13.*** Considere dos monedas cargadas. La moneda 1 tiene una probabilidad de 0.3 de caer cara, mientras que la probabilidad de la moneda 2 es de 0.6. Se tira una moneda al aire una vez; la probabilidad de que sea la moneda 1 es 0.6 y la probabilidad de que sea la moneda 2 es 0.4. El tomador de decisiones usa la regla de decisión de Bayes para determinar qué moneda se lanzó. La matriz de pagos es la siguiente:

	Estado de la naturaleza	
Alternativa	**Moneda 1 lanzada**	**Moneda 2 lanzada**
Suponga que se lanzó la moneda 1	0	−1
Suponga que se lanzo la moneda 2	−1	0
Probabilidad *a priori*	0.6	0.4

***a*)** ¿Cuál es la acción óptima antes de tirar la moneda?
***b*)** ¿Cuál es la acción óptima después de tirar la moneda si el resultado es cara? ¿Cuál si es cruz?

5.3-14 Se tienen dos monedas cargadas con probabilidades respectivas de que salga cara de 0.7 y 0.3. Se elige una de las dos monedas al azar (cada una con probabilidad $\frac{1}{2}$) para ser lanzada dos veces. El jugador recibe 250 dólares si predice correctamente cuántas caras saldrán en los dos lanzamientos.

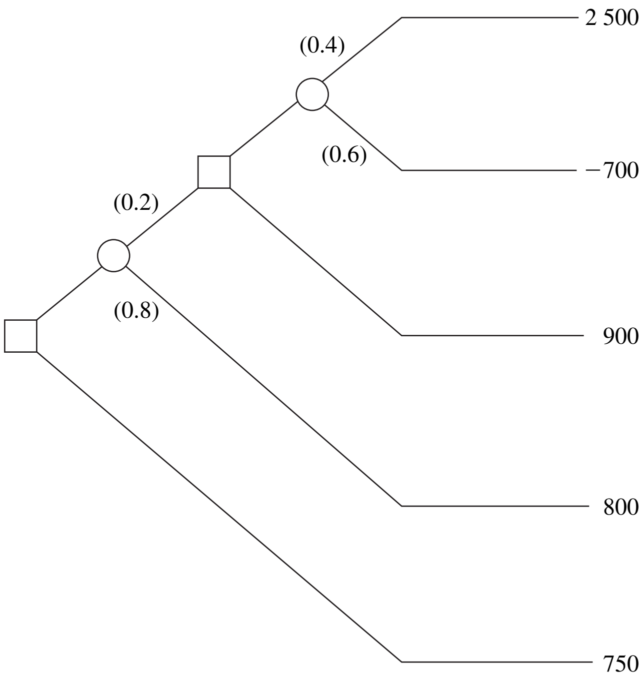
***a*)** ¿Cuál sería la predicción óptima si se usa la regla de decisión de Bayes y cuál es el pago esperado correspondiente?
T ***b*)** Suponga que ahora puede observar un tiro de prueba de la moneda elegida antes de hacer su predicción. Utilice la plantilla correspondiente de Excel para encontrar las probabilidades *a posteriori* de la moneda que se lanza.
***c*)** Determine su política óptima de predicción después de observar el tiro de prueba. ¿Cuál es el pago esperado que resulta?
***d*)** Encuentre el VEE para observar el tiro de prueba. Si debe pagar 75 dólares por observar el tiro de prueba, ¿cuál es su política óptima?

15.4-1. Lea el artículo de referencia que describe el estudio de IO que se resume en el recuadro de aplicación que se presentó en la sección 15.4. Describa de manera breve la forma en que el análisis de decisiones se aplicó a este estudio. Después, elabore una lista de los beneficios financieros y de no financieros que arrojó dicho estudio.

15.4-2.* Reconsidere el problema 15.3-2. El gerente de Silicon Dynamics desea ver el árbol de decisión del problema completo. Construya y resuelva el árbol de decisiones a mano.

15.4-3. Le proporcionan el siguiente árbol de decisión donde los números entre paréntesis son probabilidades y los números a la derecha

son los pagos en los puntos terminales. Analice el árbol de decisión para obtener la política óptima.



15.4-4.* El departamento atlético de Leland University considera la posibilidad de realizar una campaña el próximo año para reunir fondos para un nuevo campo de atletismo. En gran medida, la respuesta a la campaña depende del éxito que tenga el equipo de futbol en el otoño. En el pasado tuvo temporadas ganadoras el 60% de las veces. Si tienen una temporada ganadora (G), muchos exalumnos contribuirán y la campaña reunirá 3 millones de dólares. Si la temporada es perdedora (P), muy pocos contribuirán y perderán 2 millones de dólares. Si no se realiza la campaña, no se incurre en costo alguno. El 1 de septiembre, antes de iniciar la temporada, el departamento de atletismo debe decidir si realiza la campaña el próximo año.

a) Desarrolle una formulación de árbol de decisiones para este problema mediante la identificación de las opciones alternativas, los estados de la naturaleza y la matriz de pagos.
b) Según la regla de decisión de Bayes, ¿debe realizarse la campaña?
T ***c***) ¿Cuál es el VEIP?
d) Un famoso gurú de futbol, William Walsh, se ha ofrecido a evaluar si el equipo tendrá una temporada ganadora. Por 100 000 dólares evaluará las prácticas del equipo durante la primavera y en la pretemporada. William dará su pronóstico el 1 de septiembre respecto al tipo de temporada, G o P, que tendrá el equipo. En situaciones similares en el pasado, al evaluar equipos con temporadas ganadoras 50% del tiempo, sus predicciones fueron ciertas 75% de las veces. Si se considera que este equipo tiene una mayor tradición ganadora, si William predice una temporada ganadora, ¿cuál es la probabilidad *a posteriori* de que en realidad sea así? ¿Cuál es la probabilidad *a posteriori* de que sea perdedora? Si William predice una temporada perdedora, ¿cuál es la probabilidad *a posteriori* de que sea ganadora? ¿Y de que sea perdedora? Muestre cómo obtener las respuestas en un árbol de probabilidades.
T ***e***) Use la plantilla correspondiente de Excel para obtener las respuestas al inciso *d*).
f) Dibuje a mano el árbol de decisión para el problema completo. Analícelo para determinar la política óptima, es decir, si se debe contratar a William y si se debe realizar la campaña.

15.4-5. La contralora de Macrosoft Corporation tiene 100 millones de dólares sobrantes de fondos para inversión. Ha recibido instrucciones de invertir la cantidad completa a un año en acciones o en bonos (pero no en ambos) y después reinvertir el fondo completo ya sea en acciones o bonos (pero no en ambos) un año más. El objetivo es maximizar el valor monetario esperado del fondo al final del segundo año.

Las tasas de interés anual sobre estas inversiones dependen de la situación económica, como se muestra en la siguiente tabla:

	Tasa de interés	
Situación económica	**Acciones**	**Bonos**
Crecimiento	20%	5%
Recesión	−10%	10%
Depresión	−50%	20%

Las respectivas probabilidades de crecimiento, recesión y depresión del primer año son 0.7, 0.3 y 0. Si ocurre el crecimiento en el primer año, estas probabilidades son las mismas para el segundo año, pero, si ocurre una recesión el primer año, las probabilidades cambian a 0.2, 0.7 y 0.1 para el segundo.

a) Construya a mano el árbol de decisión del problema.
b) Analice el árbol para identificar la política óptima.

15.4-6. El lunes, cierta acción cerró a 10 dólares. El martes se espera que la acción cierre a 9, 10 u 11 dólares, con probabilidades respectivas de 0.3, 0.3 y 0.4. El miércoles, se espera que la acción cierre 10% abajo, sin cambio o 10% arriba del cierre del martes, con las siguientes probabilidades:

Cierre de hoy	10% abajo	Sin cambio	10% arriba
$ 9	0.4	0.3	0.3
$10	0.2	0.2	0.6
$11	0.1	0.2	0.7

El martes recibe instrucciones de comprar 100 acciones antes del jueves. Todas las compras se hacen al final del día, al precio de cierre conocido de ese día, de manera que sus únicas opciones son comprar al final del martes o al final del miércoles. Usted quiere determinar una estrategia óptima: comprar el martes o aplazar la compra hasta el miércoles, dado el precio al cierre del martes, con el fin de minimizar el precio esperado de compra. Desarrolle y evalúe un árbol de decisión para determinar la estrategia óptima.

15.4-7. Utilice el contexto dado en el problema 15.3-9.

a) Trace y etiquete el árbol de decisión. Incluya todos los pagos pero no las probabilidades.
T ***b***) Encuentre las probabilidades de las ramas que salen de los nodos de probabilidad.
c) Aplique el procedimiento de inducción hacia atrás e identifique la política óptima que resulta.

15.4-8. Utilice el contexto dado en el problema 15.3-11.
a) Trace y etiquete el árbol de decisión. Incluya todos los pagos pero no las probabilidades.
T *b*) Encuentre las probabilidades de las ramas que salen de los nodos de probabilidad.
c) Aplique el procedimiento de inducción hacia atrás e identifique la política óptima que resulta.

15.4-9. Utilice el contexto dado en el problema 15.3-12.
a) Dibuje y etiquete el árbol de decisión. Incluya todos los pagos pero no las probabilidades.
T *b*) Encuentre las probabilidades de las ramas que salen de los nodos de probabilidad.
c) Aplique el procedimiento de inducción hacia atrás e identifique la política óptima que resulta.

15.4-10. Use el contexto dado en el problema l5.3-13.
a) Dibuje y etiquete el árbol de decisión. Incluya todos los pagos pero no las probabilidades.
T *b*) Encuentre las probabilidades de las ramas que salen de los nodos de probabilidad.
c) Aplique el procedimiento de inducción hacia atrás e identifique la política óptima que resulta.

A **15.4-11.** La búsqueda de ejecutivos para Western Bank que realiza Headhunters Inc. está por dar frutos. El puesto que deben llenar es clave: vicepresidente de procesamiento de información, porque quien lo ocupe tendrá la responsabilidad de desarrollar un sistema de información administrativo con tecnología de punta que enlazará todas las sucursales de Western. Sin embargo, Headhunters cree haber encontrado a la persona correcta, Matthew Fenton, quien tiene excelentes referencias de un puesto similar de un banco mediano de Nueva York.

Después de una ronda de entrevistas, el presidente de Western cree que Matthew tiene una probabilidad de 0.75 de diseñar un buen sistema. Si tiene éxito, la compañía tendrá una ganancia de 4 millones de dólares (netos después del salario de Matthew, capacitación, reclutamiento y gastos). Si no tiene éxito, la compañía tendrá una pérdida de 900 000 dólares.

Por 35 000 dólares adicionales, Headhunters realizará una investigación detallada (que incluye la verificación de antecedentes, pruebas académicas y psicológicas, etc.) que precisará su potencial para el éxito. Se ha comprobado que este proceso es 90% confiable, es decir, el candidato que diseñe un buen sistema pasará la prueba con probabilidad 0.9, y el que no tendría éxito en diseñarlo fracasará con probabilidad 0.9.

La administración de Western necesita decidir si contrata a Matthew y si Headhunters debe realizar la investigación detallada antes de tomar esa decisión.
a) Construya el árbol de decisiones para este problema.
T *b*) Encuentre las probabilidades de las ramas que salen de los nodos de probabilidad.
c) Analice el árbol e identifique la política óptima.
d) Ahora suponga que la compañía puede negociar el costo de la aplicación de la prueba de aptitud a los candidatos. ¿Cuál es la cantidad máxima que debe pagar?

A **15.5-1.** Reconsidere la versión original del problema de Silicon Dinamics descrito en el problema 15.2-2.
a) Si se supone que las probabilidades *a priori* son ambas 0.5, use TreePlan para construir y resolver el árbol de decisión para este problema. De acuerdo con este análisis, ¿cuál decisión alternativa se debe elegir?
b) Use SensIt para desarrollar una gráfica que represente el pago esperado (cuando se usa la regla de decisión de Bayes) contra la probabilidad *a priori* de vender 10 000 computadoras.

A **15.5-2.** Ahora reconsidere la versión expandida del problema de Silicon Dynamics descrito en los problemas 15.3-2 y 15.4-2.
a) Use TreePlan para construir y resolver el árbol de decisión para este problema.
b) Existe cierta incertidumbre en los datos financieros ($15 millones, $6 millones y $600) establecidos en el problema 15.2-2. Cuando mucho, estas cifras podrían variar 10% de su valor base. Para cada uno de los datos realice un análisis de sensibilidad para encontrar qué podría pasar si su valor estuviera en cualquiera de los extremos de su rango de variabilidad (sin ningún cambio en los otros dos datos), mediante el ajuste de los valores en las celdas correspondientes. Después haga lo mismo para los ocho casos donde todos estos datos están en un extremo u otro de sus rangos de variabilidad.
c) Debido a la incertidumbre descrita en el inciso *b*), use SensIt para generar una gráfica que presente la ganancia esperada en el rango de variabilidad de cada dato financiero (sin ningún cambio en los otros dos datos).
d) Genere las gráficas spider y tornado correspondientes.

A **15.5-3.** Reconsidere el árbol de decisión dado en el problema 15.4-3. Use TreePlan para construir y resolver este árbol de decisión.

A **15.5-4.** Reconsidere el problema 15.4-5. Use TreePlan para construir y resolver el árbol de decisión para este problema.

A **15.5-5.** Reconsidere el problema 15.4-6. Use TreePlan para construir y resolver el árbol de decisión para este problema.

A **15.5-6.** José Morales administra un puesto de fruta en un barrio de San José, California. José compra cajas de fruta cada mañana a un agricultor al sur de San José. Cerca de 85% de las cajas resulta de buena calidad, pero el otro 15% no. Una caja buena contiene 90% de fruta excelente y dará a José una ganancia de 600 dóalres. Una caja mala contiene 40% de fruta excelente y producirá una pérdida de 2 000 dólares. Antes de que José decida aceptar la caja, tiene la opción de muestrear una pieza de fruta para probar si es o no excelente. Con base en esta muestra, tiene la opción de rechazar la caja sin pagarla. José se pregunta: 1) si debe seguir comprando a esta persona, 2) de ser así, si vale la pena muestrear sólo una pieza y 3) si lo hace, si debe aceptar o rechazar la caja con base en el resultado de la muestra.

Use TreePlan (y la plantilla de Excel para probabilidades *a posteriori*), para construir y resolver el árbol de decisión para este problema.

A **15.5-7.*** La compañía Morton Ward considera la introducción de un nuevo producto. Se piensa que tiene una posibilidad de éxito de 0.5. Una opción es probar el producto en un mercado de prueba a un costo de 5 millones de dólares, antes de tomar la decisión de introducirlo. La experiencia muestra que los productos que en última instancia tienen éxito, se aprueban en el mercado de prueba 80% de las veces, mientras que los que fracasan, el mercado de prueba los aprueba sólo 25% de las veces. Si el producto tiene éxito, la ganancia neta para la compañía será de 40 millones de dólares; si fracasa, la pérdida neta será de 15 millones de dólares.

a) Descarte la opción de probar el producto en el mercado y desarrolle una formulación de análisis de decisiones para el problema mediante la identificación de las opciones, los estados de la naturaleza y la matriz de pagos. Luego, aplique la regla de decisión de Bayes para determinar la alternativa de decisión óptima.

b) Encuentre el VEIP.

A *c*) Incluya la opción de probar el producto en el mercado, use TreePlan (y la plantilla de Excel para probabilidades *a posteriori*) para construir un árbol de decisión para este problema.

A *d*) Existe incertidumbre en las cifras de ganancias y pérdidas (40 y 15 millones de dólares). Cualquiera podría variar hasta 25% en cualquier dirección. Use TreePlan para generar una gráfica para cada una que muestre el pago esperado en este intervalo de variación.

A *e*) Debido a la incertidumbre descrita en el inciso *d*), use SensIt para generar una gráfica que muestre la ganancia esperada en el rango de variabilidad de cada una de las dos cifras financieras (sin ningún cambio en la otra cifra).

A *f*) Genere la gráfica spider y el diagrama tornado correspondientes. Interprete cada uno de ellos.

A **15.5-8.** Chelsea Bush es una candidata incipiente que busca la nominación de su partido para la presidencia de Estados Unidos. Para la elección primaria del Super Tuesday (S.T.), ella y sus consejeros piensan que si entra puede salir bien (quedar en primer o segundo lugar) o salir mal (quedar tercero o más abajo) con probabilidades respectivas de 0.4 y 0.6. Salir bien en el Super Tuesday significa que recibirá donativos cercanos a 16 millones de dólares para la campaña, mientras que si sale mal tendrá una pérdida de 10 millones de dólares después de pagar el tiempo de TV. De otra manera, puede elegir no entrar al Super Tuesday y no incurrir en gastos.

Los consejeros de Chelsea creen que sus posibilidades de éxito en el Super Tuesday pueden ser afectados por los resultados de la primaria (más pequeña) de New Hampshire (N. H.) tres semanas antes del Super Tuesday. Los analistas piensan que los resultados de la primaria de N. H. son correctos dos tercios de las veces para predecir el resultado del Super Tuesday. Entre los consejeros hay un experto en análisis de decisiones que utiliza esta información para calcular las siguientes probabilidades:

$P\{$Chelsea salga bien en S.T, dado que sale bien en N.H.$\} = \frac{4}{7}$
$P\{$Chelsea salga bien en S.T, dado que sale mal en N.H.$\} = \frac{1}{4}$
$P\{$Chelsea salga bien en N. H.$\} = \frac{7}{15}$

El costo de entrar y hacer campaña en la primaria de New Hampshire se estima en 1.6 millones de dólares.

Chelsea cree que su oportunidad de nominación depende de tener muchos fondos disponibles después del Super Tuesday para realizar una campaña vigorosa de ahí en adelante. Entonces, quiere elegir la estrategia —si debe participar en las primarias de New Hampshire y después si debe participar en las del Super Tuesday— que maximizará sus fondos esperados después de las primarias.

a) Trace y resuelva el árbol de decisión para este problema.

b) Existe cierta incertidumbre en las estimaciones de la ganancia de 16 millones de dólares o la pérdida de 10 dólares que dependen del Super Tuesday. Cualquiera de las cantidades puede diferir de la estimación hasta 25% en cualquier dirección. Para cada una de estas dos cifras financieras, realice un análisis de sensibilidad para verificar en qué medida los resultados del inciso *a*) cambiarían si el valor de la cifra financiera estuviera en cualquiera de los extremos de este rango de variabilidad (sin ningún cambio en el valor de la otra cifra). Después haga lo mismo en los cuatro casos donde ambas cifras están en alguno de los extremos de sus intervalos de variación.

A *c*) Debido a la incertidumbre descrita en el inciso *b*), use SensIt para generar una gráfica que muestre los fondos esperados por Chelsea después de estas primarias, en el intervalo de variación de cada una de las dos cifras financieras (sin ningún cambio en la otra cifra).

A *d*) Genere la gráfica spider y el diagrama tornado correspondientes. Interprete cada uno de ellos.

15.6-1. Reconsidere el ejemplo prototipo de la Goferbroke Co., que se introdujo como aplicación de utilidades en la sección 15.6. El dueño ha decidido que, dada la precaria situación económica de la empresa, necesita tomar acciones con mucha mayor aversión al riesgo. Corrigió las utilidades dadas en la tabla 15.7 de la siguiente manera: $U(-130) = 0$, $U(-100) = 0.1$, $U(60) = 0.4$, $U(90) = 0.45$, $U(670) = 0.985$ y $U(700) = 1$.

a) Analice el árbol de decisión revisado correspondiente a la figura 15.20 para obtener la nueva política óptima.

A *b*) Use TreePlan para construir y resolver el árbol de decisiones revisado.

15.6-2.* Usted vive en un área en la que se tiene la posibilidad de un temblor masivo por lo que considera comprar un seguro contra temblores para su casa a un costo anual de 180 dólares. La probabilidad de que un temblor dañe su casa en el transcurso de un año es de 0.001. Si esto ocurre, usted estima que el costo del daño (completamente cubierto por el seguro) será de 160 000 dólares. El total de sus bienes (incluso la casa) tiene un valor de 250 000 dólares.

a) Aplique la regla de decisión de Bayes para determinar la alternativa (comprar el seguro o no) que maximice el valor de sus bienes después de un año.

b) Usted desarrolló una función de utilidad que mide el valor de sus bienes en x dólares ($x \geq 0$). Esta función de utilidad es $U(x) = \sqrt{x}$. Compare la utilidad de reducir el total de sus bienes para el próximo año en un valor equivalente al del seguro, con la utilidad esperada el próximo año de no comprar el seguro de temblores. ¿Debe usted comprar el seguro?

15.6-3. Para su graduación de la universidad, sus padres le ofrecen dos alternativas. La primera es darle un regalo en dinero de 19 000 dólares. La segunda es hacer una inversión a su nombre. Esta inversión tendrá dos resultados posibles:

Resultados	Probabilidad
Recibir $10 000	0.3
Recibir $30 000	0.7

Su utilidad por recibir M dólares está dada por la función de utilidad $U(M) = \sqrt{M + 6}$. ¿Qué opción debe elegir para maximizar la utilidad esperada?

15.6-4.* Reconsidere el problema 15.6-3. Usted no está seguro de cuál es su función de utilidad para recibir el dinero, de manera que está en proceso de desarrollarla. Hasta ahora, sabe que $U(19) = 16.7$ y $U(30) = 20$ son las utilidades respectivas de recibir 19 000 y 30 000 dólares. También concluyó que es indiferente entre las dos alternativas. Use estos datos para encontrar $U(10)$.

15.6-5. Usted desea desarrollar su propia función de utilidad $U(M)$ para recibir M miles de dólares. Después de establecer $U(0) = 0$, establece $U(1) = 1$ como su utilidad por recibir 1 000 dólares. Luego desea encontrar $U(10)$ y $U(5)$.

a) Se ofrece a sí mismo las siguientes dos alternativas hipotéticas:
A_1: Obtener 10 000 dólares con probabilidad p.
Obtener 0 con una probabilidad $(1 - p)$.
A_2: En definitiva obtener 1 000 dólares.
Después se pregunta lo siguiente: ¿qué valor de p lo hace indiferente entre estas dos alternativas? Su respuesta es $p = 0.125$. Encuentre $U(10)$.

b) Después repita el inciso *a*) excepto que cambia la segunda alternativa a recibir en definitiva 5 000 dólares. El valor de p que lo hace indiferente a las dos alternativas es ahora $p = 0.5625$. Encuentre $U(5)$.

c) Repita los incisos *a*) y *b*), pero ahora use *sus* elecciones personales para p.

15.6-6. Se cuenta con la siguiente tabla de pagos:

Alternativa	Estado de la naturaleza S_1	S_2
A_1	36	49
A_2	144	0
A_3	0	81
Probabilidad *a priori*	p	$1 - p$

a) Suponga que la función de utilidad de los pagos es $U(x) = \sqrt{x}$. Grafique la utilidad esperada de cada acción alternativa contra el valor de p en la misma gráfica. Para cada alternativa, encuentre el intervalo de valores de p para el que esta alternativa maximiza la utilidad esperada.

A ***b***) Ahora suponga que su función de utilidad es la función de utilidad exponencial con una tolerancia al riesgo de $R = 50$. Use TreePlan para construir y resolver el árbol de decisión de $p = 0.25$, $p = 0.5$ y $p = 0.75$.

15.6-7. Un médico tiene un paciente seriamente enfermo con el que ha tenido problemas para diagnosticar la causa específica de su enfermedad. En este momento han disminuido las posibles causas a dos: enfermedad A o enfermedad B. Según la evidencia con que cuenta, piensa que las dos alternativas tienen iguales probabilidades de ser correctas.

Ya no dispone de más pruebas que las que realizó para determinar si la causa es la enfermedad B. Puede hacer un análisis de la enfermedad A, pero tiene dos problemas. Primero, es muy costoso. Segundo, es poco confiable pues da resultados acertados sólo 80% de las veces; esto es, da resultados positivos (que indican la enfermedad A) sólo para 80% de los pacientes que la padecen, y da resultados positivos para 20% de los pacientes que en realidad tienen la enfermedad B.

La enfermedad B es muy seria y no se conoce un tratamiento. Algunas veces es fatal y los que sobreviven quedan con mala salud y mala calidad de vida. El diagnóstico es similar para los pacientes de la enfermedad A si se deja sin tratamiento, pero se dispone de un tratamiento costoso que elimina el peligro de muerte a quienes la padecen, y que puede regresarles su buena salud. Sin embargo, es un tratamiento radical que siempre conduce a la muerte si los pacientes tienen la enfermedad B.

La distribución de probabilidad del diagnóstico del paciente en cuestión está dada para cada caso en la siguiente tabla, donde los títulos de las columnas (después de la primera) indican la enfermedad del paciente.

	Probabilidades de los resultados			
	Sin tratamiento		Con tratamiento para la enfermedad *A*	
Resultado	***A***	***B***	***A***	***B***
Muere	0.2	0.5	0	1.0
Sobrevive con mala salud	0.8	0.5	0.5	0
Recupera la buena salud	0	0	0.5	0

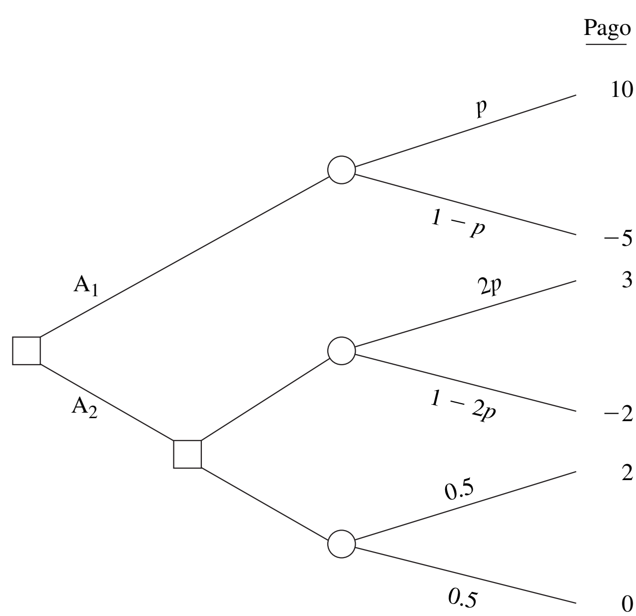
El paciente ha asignado las siguientes utilidades a los resultados posibles:

Resultado	Utilidad
Muere	0
Sobrevive con mala salud	10
Recupera la buena salud	30

Además, estas utilidades deben incrementarse en –2 si el paciente incurre en el costo de las pruebas de la enfermedad A y en –1 si incurre en el costo del tratamiento de la enfermedad A.

Utilice análisis de decisiones con un árbol de decisión completo para determinar si el paciente debe aceptar que le hagan la prueba de la enfermedad A y cómo debe proceder después (recibir o no tratamiento para la enfermedad A) para maximizar la utilidad esperada del paciente.

15.6-8. Se quiere elegir entre las acciones A_1 y A_2 que se muestran en el siguiente árbol de decisiones, pero existe incertidumbre respecto al valor de la probabilidad p, por lo que debe realizar un análisis de sensibilidad de p.



La función de utilidad del dinero (pago recibido) es

$$U(M) = \begin{cases} M^2 & \text{si } M \geq 0 \\ M & \text{si } M < 0. \end{cases}$$

a) Para $p = 0.25$, determine qué acción es óptima para maximizar la utilidad esperada del pago.

b) Determine el intervalo de valores de la probabilidad p ($0 \leq p \leq 0.5$) para el que la misma acción permanece óptima.

CASOS

Caso 15.1 Negocios inteligentes

Mientras El Niño derrama lluvia en el norte de California, Charlotte Rothstein, directora ejecutiva, accionista mayoritaria y fundadora de Cerebrosoft, sentada en su oficina contempla la decisión que debe tomar respecto del producto propuesto más nuevo de su compañía, Brainet. Ésta es una decisión en particular difícil. Brainet puede captar la atención y venderse bien. Sin embargo, Charlotte está preocupada por el riesgo que implica. En este mercado competitivo, su comercialización también podría provocar pérdidas sustanciales. ¿Debe seguir adelante e iniciar la campaña de ventas? ¿Debe abandonar el producto? ¿Debe quizá comprar información de investigación del mercado adicional a una compañía local que se dedica a esto antes de decidir lanzar el producto? Tiene que tomar la decisión muy pronto, y por eso, mientras bebe despacio su jugo multivitamínico alto en proteínas, reflexiona sobre los eventos de los últimos años.

Cerebrosoft fue fundada por Charlotte y dos amigos después de salir de la carrera de administración. La compañía se localiza en el corazón de Silicon Valley. Charlotte y sus amigos se las arreglaron para ganar dinero en el segundo año de operación y siguen ganándolo. Cerebrosoft fue una de las primeras compañías que vendió software en internet y desarrolló herramientas para PC para el sector de multimedia. Dos de los productos generan 80% de los ingresos de la compañía, Audiatur y Videatur. Se vendieron más de 100 000 unidades de cada uno durante el año pasado. El negocio se hace en internet: los clientes pueden bajar una versión de prueba del software, probarlo y, si están satisfechos con lo que ven, pueden comprarlo (por medio de una clave que les permite desactivar el contador de tiempo en la versión de prueba.) Ambos productos tienen un precio de $75.95 y sólo se venden en internet.

Aunque internet es una red de computadoras de diferentes tipos, que corren tipos distintos de software, un protocolo estandarizado les permite comunicarse. Los usuarios pueden "navegar" en la red y visitar computadoras que se encuentran a miles de millas y tener acceso a la información disponible en el sitio. También pueden tener archivos disponibles en internet y así es como Cerebrosoft genera sus ventas. La venta del software en internet elimina muchos factores de costo tradicionales de los productos para el consumidor, como empaque, almacenamiento, distribución, fuerza de ventas, etc. En su lugar, los clientes potenciales: pueden bajar la versión de prueba y ver el producto (es decir, usarlo) antes de que expire el periodo de prueba, y después decidir si lo quieren comprar. Además, Cerebrosoft puede poner a disposición del cliente los archivos más recientes, con lo que evita el problema de tener software obsoleto en la línea de distribución.

La llegada de Jeannie Korn interrumpe los pensamientos de Charlotte. Ella está a cargo del marketing de una línea de productos y Brainet ha tenido su atención desde el principio. Está lista para dar a Charlotte la opinión que pidió: "Charlotte, de verdad creo que debes seguir adelante con Brainet. Los ingenieros de software me convencieron de que la versión actual es poderosa y ¡queremos estar en el mercado con esto lo antes posible! De los datos de lanzamiento de nuestros productos de los últimos dos años podemos obtener una estimación bastante confiable de cómo responderá el mercado al nuevo producto, ¿no crees? Además, mira". Saca unas diapositivas de presentación: "Durante ese tiempo hemos lanzado 12 nuevos productos en total y 4 de ellos han tenido una venta de más de 30 000 unidades nada más en los primeros seis meses. Todavía mejor: los dos últimos que se lanzaron vendieron más de 40 000 copias durante los dos primeros trimestres". Charlotte conoce estos números tan bien como Jeannie. Después de todo, dos de estos lanzamientos han sido productos que ella misma ayudó a desarrollar. Pero no se siente tranquila con este lanzamiento específico. La compañía ha crecido con rapidez en los últimos tres años y su capacidad financiera se encuentra cerca del límite, un lanzamiento malo del Brainet le costaría a la compañía mucho dinero, algo que no está disponible por ahora debido a las inversiones que Cerebrosoft ha hecho.

Después, esa tarde, Charlotte se reúne con Reggie Ruffin, capaz en todas las áreas y gerente de producción. Reggie tiene una trayectoria sólida en su campo y Charlotte quiere su opinión sobre el proyecto Brainet.

"Bueno, Charlotte, con mucha franqueza pienso que existen tres factores principales relevantes para el éxito de este proyecto: competencia, unidades vendidas, ah, y por supuesto, el precio. ¿Ya decidiste el precio?"

"Todavía estoy estudiando cuál de tres estrategias nos daría el mayor beneficio. Vender en 50 dólares y tratar de maximizar las ganancias, o vender en 30 dólares y tratar de maximizar el porcentaje de mercado. Por supuesto, resta tu tercera alternativa, vender en 40 dólares y tratar de lograr las dos cosas."

En este punto Reggie pone atención en la hoja de papel frente a él. "Yo pienso que la alternativa de 40 dólares es la mejor. Respecto de los costos, verifiqué los registros: en principio, tenemos que amortizar los costos de desarrollo en que incurrimos para Brainet. Hasta ahora se han gastado 800 000 dólares y esperamos gastar 50 000 dólares por año en apoyo y envío de CD a quienes quieran su copia además de la que bajaron". Después Reggie pasa un informe a Charlotte. "Aquí tenemos algunos datos de la industria. Ayer recibí esto, directo de la prensa. Veamos lo que podemos averiguar de la industria." Muestra a Charlotte algunos puntos importantes, Reggie accede

a resumir la información más relevante del informe y tenerlo listo la mañana siguiente para Charlotte. Trabajó hasta muy tarde en la noche para reunir los datos de las páginas del informe, pero al final obtuvo tres tablas, una para cada alternativa de las estrategias de precios. Cada tabla muestra la probabilidad correspondiente de varias cantidades de ventas dado el nivel de competencia (alto, medio o bajo) que desarrollen las otras compañías.

La mañana siguiente Charlotte bebe otra de sus bebidas energéticas. Jeannie y Reggie llegarán a su oficina en cualquier momento y, con su ayuda, tendrá que decidir qué hacer con Brainet. ¿Debe lanzar el producto? Si así es, ¿qué precio debe tener?

Cuando Jeannie y Reggie entran en la oficina, Jeannie exclama: "Amigos, acabo de hablar con nuestra compañía de estudio de mercado. Dicen que pueden hacer el estudio sobre la situación de la competencia para la introducción de Brainet y entregar los resultados en una semana."

"¿Cuánto quieren por el estudio?"

"Sabía que me preguntarías eso, Reggie. Quieren 10 000 dólares, y pienso que es un precio justo."

En este punto Charlotte interviene en la conversación. "¿Tenemos algunos datos acerca de la calidad del trabajo de está compañía?"

"Sí, yo tengo unos informes aquí. Después de analizarlos, llegué a la conclusión de que la compañía de estudios de mercado no es muy buena para predecir el entorno competitivo para precios medios o bajos. Por lo tanto, no debemos pedirles que hagan el estudio si nos decidimos por una de estas estrategias de precio. Pero en el caso de un precio alto, lo hacen bastante bien: dado que la competencia resulta ser alta, ellos han proporcionado el pronóstico correcto 80% del tiempo, mientras que 15% del tiempo pronosticaron competencia media en esa situación. Cuando la competencia resulta ser media, han pronosticado competencia alta 15% del tiempo y media 80% del tiempo. Por último, en el caso de competencia baja, los números fueron 90% de las veces un pronóstico correcto, 7% una predicción 'media' y 3% una predicción 'alta'".

Charlotte cree que estos números son demasiado para ella. "¿No contamos con una estimación sencilla de cómo reaccionará el mercado?"

"¿Te refieres a probabilidades *a priori*? Seguro, por nuestra experiencia, la posibilidad de enfrentar una competencia alta es de 20%, mientras que es de 70% para una competencia media y 10% para una baja". Jeannie siempre tiene las cifras listas cuando se necesitan.

Ahora sólo queda sentarse y darle algún sentido a todo esto…

TABLA 1 Distribución de probabilidad de unidades vendidas dado un precio alto ($50)

	Nivel de competencia		
Ventas	**Alto**	**Medio**	**Bajo**
50 000 unidades	0.2	0.25	0.3
30 000 unidades	0.25	0.3	0.35
20 000 unidades	0.55	0.45	0.35

TABLA 2 Distribución de probabilidad de unidades vendidas, dado un precio medio ($40)

	Nivel de competencia		
Ventas	**Alto**	**Medio**	**Bajo**
50 000 unidades	0.25	0.30	0.40
30 000 unidades	0.35	0.40	0.50
20 000 unidades	0.40	0.30	0.10

TABLA 3 Distribución de probabilidad de unidades vendidas, dado un precio bajo ($30)

	Nivel de competencia		
Ventas	**Alto**	**Medio**	**Bajo**
50 000 unidades	0.35	0.40	0.50
30 000 unidades	0.40	0.50	0.45
20 000 unidades	0.25	0.10	0.05

a) Para el análisis inicial pase por alto la oportunidad de obtener más información con la contratación de la compañía de estudios de mercado. Identifique las acciones alternativas y los estados de la naturaleza. Construya la matriz de pagos. Después formule el problema de decisión como un árbol de decisiones. Distinga con claridad entre los nodos de decisión y los de probabilidad e incluya los datos relevantes.

b) ¿Cuál debe ser la decisión de Charlotte si usa el criterio de la posibilidad máxima? ¿El criterio del pago máximo?

c) ¿Cuál es la decisión de Charlotte si usa la regla de decisión de Bayes?

d) Ahora considere la posibilidad de realizar el estudio de mercado. Desarrolle el árbol de decisión correspondiente. Calcule las probabilidades relevantes y analice el árbol de decisión. ¿Debe Cerebrosoft pagar 10 000 dólares por la investigación de mercado? ¿Cuál es la política óptima global?

RESUMEN DE LOS CASOS ADICIONALES EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

CASO 15.2 Apoyo inteligente al conductor

El director ejecutivo de Bay Area Automobile Gadgets contempla la posibilidad de agregar un dispositivo examinador de caminos al sistema de soporte al conductor de la compañía. Se debe tomar una serie de decisiones. ¿Es necesario realizar investigación básica en el examinador de caminos? Si la investigación es exitosa, ¿debe la compañía desarrollar el producto o vender la tecnología? En el caso de un desarrollo exitoso del producto, ¿debería la compañía lanzar el producto al mercado o vender el concepto? Es necesario aplicar análisis de decisiones para afrontar estos temas. Parte del análisis involucrará el uso de la función de utilidad del director ejecutivo.

CASO 15.3 ¿Quién quiere ser millonario?

Usted es un participante en el juego "¿Quién quiere ser millonario?" y acaba de responder correctamente la pregunta de los 250 000 dólares. Si usted decide seguir adelante con la pregunta de los 500 000 dólares y después con la del millón del dólares, tiene la opción de usar la línea "telefonee a un amigo" en una de las preguntas con el fin de mejorar sus probabilidades de contestar correctamente. Usted desea usar el análisis de decisiones (incluyendo el árbol de decisiones y la teoría de la utilidad) para decidir cómo proceder.

CASO 15.4 University Toys y los personajes de acción de un profesor de ingeniería

University Toys ha desarrollado una serie de personajes de acción de un profesor de ingeniería para la escuela de ingeniería local y la administración debe decidir cómo vender los muñecos en un ambiente de incertidumbre acerca de la demanda de este producto. Una opción consiste en tomar la decisión de invertir todo el esfuerzo en su producción, publicidad y ventas. Otra, consiste en probar el producto en el mercado en primera instancia. Un problema con esta última opción es que hay un rumor de que la competencia está por lanzar al mercado un producto similar. Es necesario llevar a cabo el análisis de decisiones (incluyendo el árbol de decisiones y el análisis de sensibilidad) con el fin de decidir la manera de proceder.

16

CAPÍTULO

Cadenas de Markov

El capítulo 15 se enfocó en la toma de decisiones ante la incertidumbre de *uno* o más eventos futuros, con la intención de conocer el verdadero estado de la naturaleza. Sin embargo, algunas decisiones deben tomar en cuenta la incertidumbre acerca de *muchos* eventos futuros. Ahora se presentará el fundamento de la toma de decisiones en este contexto más amplio.

En particular, este capítulo presenta modelos de probabilidad de procesos que *evolucionan en el tiempo* de una manera probabilística. Tales procesos se llaman *procesos estocásticos*. Después de una introducción breve de los procesos estocásticos generales en la primera sección, el resto del capítulo se dedica a un tipo especial de proceso que se denomina *cadena de Markov*. Las cadenas de Markov tienen la propiedad particular de que las probabilidades que describen la forma en que el proceso evolucionará en el futuro dependen sólo del estado actual en que se encuentra el proceso y, por lo tanto, son independientes de los eventos que ocurrieron en el pasado. Muchos procesos se ajustan a esta descripción, por lo que las cadenas de Markov constituyen una clase de modelo probabilístico de gran importancia.

Por ejemplo, usted podrá observar en el capítulo siguiente que las *cadenas de Markov de tiempo continuo* (descritas en la sección 16.8) se utilizan para formular gran parte de los modelos básicos de la *teoría de colas*. Las cadenas de Markov también proporcionan las bases para el estudio de los *modelos de decisión de Markov* en el capítulo 19. Existe una amplia gama de aplicaciones de las cadenas de Markov. Un gran número de libros y artículos presentan algunas de estas aplicaciones. Una de ellas constituye la referencia seleccionada 4 que describe las aplicaciones en áreas tan diversas como la clasificación de los clientes, la secuencia del DNA, el análisis de redes genéticas, la estimación de la demanda de ventas a través del tiempo y el valor del crédito. También se presenta, en el recuadro de aplicación de la sección 16.2, un ejemplo que involucra el valor del crédito, mientras que el recuadro de aplicación de la sección 16.8 se relaciona con el mantenimiento de máquinas. La referencia seleccionada 6 hace hincapié en las aplicaciones en finanzas y la 3 describe aplicaciones para el análisis de la estrategia del beisbol. La lista continúa de manera indefinida, pero por el momento veremos una descripción de los procesos estocásticos en general y de las cadenas de Markov en particular.

16.1 PROCESOS ESTOCÁSTICOS

Un **proceso estocástico** se define como una colección indexada de variables aleatorias $\{X_t\}$, donde el índice t toma valores de un conjunto T dado. Con frecuencia T se considera el conjunto de enteros no negativos mientras que X_t representa una característica de interés cuantificable en el tiempo t. Por ejemplo, X_t puede representar los niveles de inventario al final de la semana t.

Los procesos estocásticos son de interés para describir el comportamiento de un sistema en operación durante algunos periodos. Un proceso estocástico tiene la siguiente estructura.

> La condición actual del sistema puede estar en una de $M + 1$ categorías mutuamente excluyentes llamadas **estados**. Por conveniencia en la notación, estos estados se etiquetan $0, 1, 2, \ldots, M$. La

variable aleatoria X_t representa el *estado del sistema* en el tiempo t, de manera que sus únicos valores posibles son 0, 1, . . ., M. El sistema se observa en puntos del tiempo dados, etiquetados $t = 0, 1, 2, \ldots$ De esta forma, los procesos estocásticos $\{X_t\} = \{X_0, X_1, X_2, \ldots\}$ proporcionan una representación matemática de la forma en que evoluciona la condición del sistema físico a través del tiempo.

Este tipo de procesos se conocen como procesos estocásticos de *tiempo discreto* con *espacio de estados finito*. Excepto en el caso de la sección 16.8, éste será el único tipo de procesos estocásticos que se estudiará en este capítulo. (En la sección 16.8 se describen ciertos procesos estocásticos de *tiempo continuo*.)

Ejemplo de clima

El clima en el pueblo de Centerville puede cambiar con rapidez de un día a otro. Sin embargo, las posibilidades de tener clima seco (sin lluvia) mañana es de alguna forma mayor si hoy está seco, es decir, si no llueve. En particular, la probabilidad de que mañana esté seco es de **0.8** si hoy está seco, pero es de sólo **0.6** si hoy llueve. Estas probabilidades no cambian si se considera la información acerca del clima en los días anteriores a hoy.

La evolución del clima día tras día en Centerville es un proceso estocástico. Si se comienza en algún día inicial (etiquetado como día 0), el clima se observa cada día t, para $t = 0, 1, 2, \ldots$ El estado del sistema en el día t puede ser

Estado 0 = El día t es seco

o bien

Estado 1 = El día t es lluvioso

Así, para $t = 0, 1, 2, \ldots$, la variable aleatoria X_t toma los valores,

$$X_t = \begin{cases} 0 \text{ si día } t \text{ es seco} \\ 1 \text{ si día } t \text{ es lluvioso} \end{cases}$$

El proceso estocástico $\{X_t\} = \{X_0, X_1, X_2, \ldots\}$ proporciona una representación matemática de la forma en que evoluciona el clima en Centerville a través del tiempo.

Ejemplo de inventarios

La tienda de fotografía de Dave tiene el siguiente problema de inventario. El negocio tiene en almacén un modelo especial de cámara que se puede solicitar cada semana. Sean $D_1, D_2, \ldots$ representan las *demandas* respectivas de esta cámara (el número de unidades que se venderían si el inventario no se agota) durante la primera, segunda semanas, . . ., respectivamente, entonces, la variable aleatoria D_t (para $t = 1, 2, \ldots$) es

D_t = número de cámaras que se venderían en la semana t si el inventario no se agota. (Este número incluye las ventas perdidas cuando se agota el inventario.)

Se supone que las D_t son variables aleatorias independientes e idénticamente distribuidas que tienen una *distribución Poisson* con media de 1. Sea X_0 el número de cámaras que se tiene en el momento de iniciar el proceso, X_1 el número de cámaras que se tienen al final de la semana 1, X_2 el número de cámaras al final de la semana 2, etc., entonces la variable aleatoria X_t (para $t = 0, 1, 2, \ldots$) es

X_t = número de cámaras disponibles al final de la semana t.

Suponga que $X_0 = 3$, de manera que la semana 1 se inicia con tres cámaras a mano.

$$\{X_t\} = \{X_0, X_1, X_2, \ldots\}$$

es un proceso estocástico donde la variable aleatoria X_t representa el estado del sistema en el tiempo t, a saber

Estado en el tiempo t = número de cámaras disponibles al final de la semana t.

Como propietario de la tienda, Dave desearía aprender más acerca de cómo evoluciona este proceso estocástico a través del tiempo mientras se utilice la política de pedidos actual que se describe a continuación.

Al final de cada semana t (sábado en la noche), la tienda hace un pedido que le entregan en el momento de abrir la tienda el lunes. La tienda usa la siguiente política para ordenar:

Si $X_t = 0$, ordena 3 cámaras.
Si $X_t > 0$, no ordena ninguna cámara.

En consecuencia, el nivel de inventarios fluctúa entre un mínimo de cero y un máximo de tres cámaras, por lo que los estados posibles del sistema en el tiempo t (al final de la semana t) son

Estados posibles = 0, 1, 2, o 3 cámaras disponibles.

Como cada variable aleatoria X_t (t = 0, 1, 2, . . .) representa el estado del sistema al final de la semana t, sus únicos valores posibles son 0, 1, 2, 3. Las variables aleatorias X_t son dependientes y se pueden evaluar en forma iterativa por medio de la expresión

$$X_t+_1 \begin{cases} \text{máx}\{3 - D_{t+1}, 0\} & \text{si} \quad X_t = 0 \\ \text{máx}\{X_t - D_{t+1}, 0\} & \text{si} \quad X_t \geq 1, \end{cases}$$

para t = 0, 1, 2, . . .

Estos ejemplos se usarán como ilustración en muchas de las secciones que siguen. En la sección 16.2 se define con más detalle el tipo de procesos estocásticos que se analizará en este capítulo.

16.2 CADENAS DE MARKOV

Es necesario hacer algunos supuestos sobre la distribución conjunta de X_0, X_1, . . . para obtener resultados analíticos. Un supuesto que conduce al manejo analítico es que el proceso estocástico es una cadena de Markov, que tiene la siguiente propiedad esencial:

Se dice que un proceso estocástico $\{X_t\}$ tiene la **propiedad markoviana** si $P\{X_{t+1} = j|X_0 = k_0, X_1 = k_1, \ldots, X_{t-1} = k_{t-1}, X_t = i\} = P\{X_{t+1} = j|X_t = i\}$, para t = 0, 1, . . . y toda sucesión $i, j, k_0, k_1, \ldots, k_{t-1}$.

En palabras, esta propiedad markoviana establece que la probabilidad condicional de cualquier "evento" futuro dados cualquier "evento" pasado y el estado actual $X_t = i$, es *independiente* de los eventos pasados y sólo depende del estado actual del proceso.

Un proceso estocástico $\{X_t\}$ (t = 0, 1, . . .) es una **cadena de Markov** si presenta la *propiedad markoviana*.

Las probabilidades condicionales $P\{X_{t+1} = j|X_t = i\}$ de una cadena de Markov se llaman **probabilidades de transición** (de un paso). Si para cada i y j,

$$P\{X_{t+1} = j \mid X_t = i\} = P\{X_1 = j \mid X_0 = i\}, \quad \text{para toda } t = 1, 2, \ldots,$$

entonces se dice que las probabilidades de transición (de un paso) son *estacionarias*. Así, tener **probabilidades de transición estacionarias** implica que las probabilidades de transición no cambian con el tiempo. La existencia de probabilidades de transición (de un paso) estacionarias también implica que, para cada i, j y n (n = 0, 1, 2, . . .),

$$P\{X_{t+n} = j \mid X_t = i\} = P\{X_n = j \mid X_0 = i\}$$

para toda t = 0, 1, . . . Estas probabilidades condicionales se llaman **probabilidades de transición de n pasos**.

Para simplificar la notación de las probabilidades de transición estacionarias, sea

$$p_{ij} = P\{X_{t+1} = j | X_t = i\},$$
$$p_{ij}^{(n)} = P\{X_{t+n} = j | X_t = i\}.$$

Así, las probabilidades de transición de n pasos $p_{ij}^{(n)}$ son simplemente la probabilidad condicional de que el sistema se encuentre en el estado j exactamente después de n pasos (unidades de tiempo), dado que comenzó en el estado i en cualquier tiempo t. Cuando $n = 1$ observe que $p_{ij}^{(1)} = p_{ij}$.[1]

Como las $p_{ij}^{(n)}$ son probabilidades condicionales, deben ser no negativas y, como el proceso debe hacer una transición a algún estado, deben satisfacer las propiedades

$$p_{ij}^{(n)} \geq 0, \qquad \text{para toda } i \text{ y } j;\ n = 0, 1, 2, \ldots,$$

y

$$\sum_{j=0}^{M} p_{ij}^{(n)} = 1 \qquad \text{para toda } i;\ n = 0, 1, 2, \ldots$$

Una notación conveniente para representar las probabilidades de transición de n pasos es la *matriz de transición de n pasos*

$$\mathbf{P}^{(n)} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ \vdots \\ M \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & \cdots & M \end{array} \\ \begin{bmatrix} p_{00}^{(n)} & p_{01}^{(n)} & \cdots & p_{0M}^{(n)} \\ p_{10}^{(n)} & p_{11}^{(n)} & \cdots & p_{1M}^{(n)} \\ \cdots & \cdots & \cdots & \cdots \\ p_{M0}^{(n)} & p_{M1}^{(n)} & & p_{MM}^{(n)} \end{bmatrix} \end{array}$$

Observe que la probabilidad de transición en un renglón y columna dados es la de la transición *del* estado en ese renglón *al* estado en la columna. Cuando $n = 1$, el superíndice n no se escribe y se hace referencia a ésta como una *matriz de transición*.

Las cadenas de Markov que se estudian en este capítulo tienen las siguientes propiedades:

1. Un número finito de estados.
2. Probabilidades de transición estacionarias.

También se supondrá que se conocen las probabilidades iniciales $P\{X_0 = i\}$ para toda i.

Formulación del ejemplo del clima como una cadena de Markov

En el ejemplo del clima que se presentó en la sección anterior, recuerde que la evolución del clima día tras día en Centerville se ha formulado como un proceso estocástico $\{X_t\}$ $(t = 0, 1, 2, \ldots)$ donde

$$X_t = \begin{cases} 0 \text{ si día } t \text{ es seco} \\ 1 \text{ si día } t \text{ es lluvioso} \end{cases}$$

$$P\{X_{t+1} = 0 | X_t = 0\} = 0.8,$$
$$P\{X_{t+1} = 0 | X_t = 1\} = 0.6$$

Aún más, como estas probabilidades no cambian si también se toma en cuenta la información del clima antes del día de hoy (día t),

$$P\{X_{t+1} = 0 | X_0 = k_0, X_1 = k_1, \ldots, X_{t-1} = k_{t-1}, X_t = 0\} = P\{X_{t+1} = 0 | X_t = 0\}$$
$$P\{X_{t+1} = 0 | X_0 = k_0, X_1 = k_1, \ldots, X_{t-1} = k_{t-1}, X_t = 1\} = P\{X_{t+1} = 0 | X_t = 1\}$$

[1] Para $n = 0$, $p_{ij}^{(0)}$ es $P\{X_0 = j | X_0 = i\}$, que es igual a 1 cuando $i = j$ y es 0 cuando $i \neq j$.

Recuadro de aplicación

Merrill Lynch es una firma líder en servicios integrales financieros que brinda servicios de corretaje, de inversiones y bancarios a clientes al menudeo y a pequeños negocios a la vez que ayuda a las grandes corporaciones e instituciones alrededor del mundo a formar capital. Uno de los afiliados de Merrill Lynch, *Merrill Lynch (ML) Bank USA*, cuenta con activos por más de 60 000 millones de dólares obtenidos mediante la aceptación de depósitos provenientes de los clientes al menudeo de Merrill Lynch a los que utiliza para financiar préstamos y realizar inversiones.

En el año 2000, ML Bank USA comenzó a otorgar líneas de crédito revolvente a sus clientes corporativos. En cosa de unos años, el banco había desarrollado un portafolios de alrededor de 13 000 millones de dólares en compromisos de líneas de crédito con más de 100 instituciones. Antes de llegar a este punto, se le solicitó al eficiente grupo de investigación de operaciones guiar a los administradores de éste cada vez más grande portafolios mediante el uso de técnicas de investigación de operaciones con el fin de evaluar el *riesgo de liquidez* (situación que consiste en que el banco no pueda cumplir con sus obligaciones de dinero en efectivo) asociado con sus obligaciones de línea de crédito actuales y en el futuro.

El grupo de investigación de operaciones desarrolló un *modelo de simulación* (el tema del capítulo 20) con este objetivo. Sin embargo, el más importante potencial de este modelo consiste en una *cadena de Markov* que describe la evolución del valor del crédito de cada cliente a través del tiempo. Los estados de la cadena de Markov son los diferentes valores posibles del crédito (que varían desde *la inversión más alta hasta la inversión por omisión*) que se le asignan a las principales compañías a través de agencias de clasificación de créditos como Standard and Poor's y Moody's. La probabilidad de transición desde el estado i hasta el j de la matriz de transición de una determinada compañía se define como la probabilidad de que la agencia de clasificación de créditos cambie la clasificación de la compañía del estado i al j de un mes a otro con base en los patrones históricos de compañías similares.

Esta aplicación de la investigación de operaciones (incluyendo las cadenas de Markov) permitió a ML Bank USA *liberar alrededor de* **4 000 millones de dólares** *de liquidez para utilizarlos en otra cosa*, así como también expandir su portafolios de compromisos en líneas de crédito más de 60% en menos de dos años. Otros de los beneficios que se obtuvieron a partir de dicho estudio fueron la capacidad para evaluar escenarios muy riesgosos y llevar a cabo la planeación a largo plazo. Este sorprendente trabajo hizo que Merril Lynch se hiciera merecedora del prestigiado *Premio Wagner para la Excelencia en la Práctica de la Investigación de Operaciones* en 2004.

Fuente: Duffy, T., M. Hatzakis, W. Hsu, R. Labe, B. Liao, X. Luo, J. Oh, A. Setya y L. Yang: "Merrill Lynch Improves Liquidity Risk Management for Revolving Credit Lines", en *Interfaces*, **35**(5): 353-369, septiembre-octubre de 2005. (En el sitio en internet de este libro —www.mhhe.com/hillier— se proporciona una liga hacia este artículo.)

para $t = 0, 1, \ldots$ y toda sucesión $k_0, k_1, \ldots, k_{t-1}$. Estas ecuaciones también deben cumplirse si $X_{t+1} = 0$ se reemplaza con $X_{t+1} = 1$. (La razón es que los estados 0 y 1 son mutuamente excluyentes y son los únicos estados posibles; por ende, las probabilidades de los dos estados deben sumar 1.) Por lo tanto, el proceso estocástico tiene la *propiedad markoviana*, lo que lo convierte en una cadena de Markov.

Si se usa la notación que se introdujo en esta sección, las probabilidades de transición (de un paso) son

$$p_{00} = P\{X_{t+1} = 0 \mid X_t = 0\} = 0.8,$$
$$p_{10} = P\{X_{t+1} = 0 \mid X_t = 1\} = 0.6$$

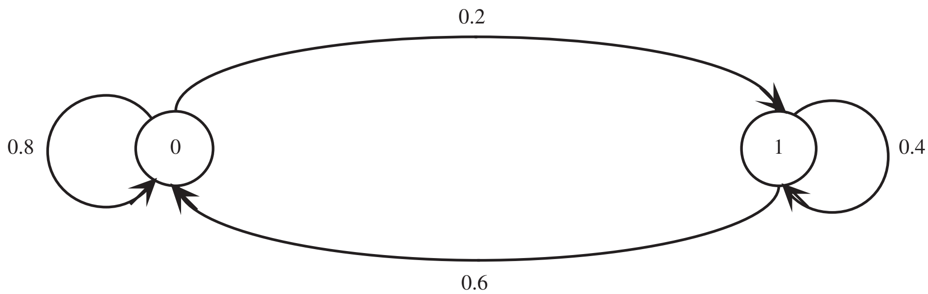
para toda $t = 1, 2, \ldots$, por lo que éstas son las probabilidades de transición *estacionarias*. Además,

$$p_{00} + p_{01} = 1 \quad \text{entonces} \quad p_{01} = 1 - 0.8 = 0.2,$$
$$p_{10} + p_{11} = 1 \quad \text{entonces} \quad p_{11} = 1 - 0.6 = 0.4,$$

Por lo tanto, la matriz de transición es

$$\mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} \end{array} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} \end{array}$$

donde estas probabilidades de transición se refieren a la transición *del* estado del renglón *al* estado de la columna. Tenga en mente que el estado 0 hace referencia a un día seco, mientras que el estado 1 significa que el día es lluvioso, así que estas probabilidades de transición proporcionan la probabilidad del estado del clima el día de mañana, dado el estado del clima del día de hoy.



■ **FIGURA 16.1**
Diagrama de transición de estados del ejemplo del clima.

El diagrama de transición de estado de la figura 16.1 presenta de manera gráfica la misma información que proporciona la matriz de transición. Los nodos (círculos) representan los dos estados posibles del clima, mientras que las flechas muestran las transiciones posibles de un día al siguiente (se incluye la opción de regresar al mismo estado). Cada una de las probabilidades de transición se escribe a continuación de la flecha correspondiente.

Las matrices de transición de n pasos se presentarán en la siguiente sección.

Formulación del ejemplo de inventarios como una cadena de Markov

Si se regresa al ejemplo de inventarios que se desarrolló en la sección anterior, recuerde que X_t es el número de cámaras en almacén al final de la semana t (antes de ordenar más), donde X_t representa el *estado del sistema* en el tiempo t (el fin de la semana t). Dado que el estado actual es $X_t = i$, la expresión al final de la sección 16.1 indica que X_{t+1} depende sólo de D_{t+1} (la demanda en la semana $t + 1$) y X_t. Como X_{t+1} es independiente de la historia del sistema de inventarios antes del tiempo t, el proceso estocástico $\{X_t\}$ $(t = 0, 1, \ldots)$ tiene la *propiedad markoviana* y por lo tanto es una cadena de Markov.

A continuación se verá cómo obtener las probabilidades de transición (de un paso), es decir, los elementos de la *matriz de transición* (de un paso)

$$\mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} \\ p_{10} & p_{11} & p_{12} & p_{13} \\ p_{20} & p_{21} & p_{22} & p_{23} \\ p_{30} & p_{31} & p_{32} & p_{33} \end{bmatrix} \end{array}$$

dado que D_{t+1} tiene una distribución Poisson con media de 1. Entonces,

$$P\{D_{t+1} = n\} = \frac{(1)^n e^{-1}}{n!}, \qquad \text{para } n = 0, 1, \ldots,$$

por lo que (expresando con tres dígitos significativos) se tiene,

$$\begin{aligned} &P\{D_{t+1} = 0\} = e^{-1} = 0.368, \\ &P\{D_{t+1} = 1\} = e^{-1} = 0.368, \\ &P\{D_{t+1} = 2\} = \frac{1}{2}e^{-1} = 0.184, \\ &P\{D_{t+1} \geq 3\} = 1 - P\{D_{t+1} \leq 2\} = 1 - (0.368 + 0.368 + 0.184) = 0.080. \end{aligned}$$

Para el primer renglón de **P** se trata de la transición del estado $X_t = 0$ a algún estado X_{t+1}. Como se indica al final de la sección 16.1,

$$X_{t+1} = \text{máx}\{3 - D_{t+1}, 0\} \quad \text{si} \quad X_t = 0.$$

Por lo tanto, para la transición a $X_{t+1} = 3$, o $X_{t+1} = 2$ o $X_{t+1} = 1$,

$$p_{03} = P\{D_{t+1} = 0\} = 0.368,$$
$$p_{02} = P\{D_{t+1} = 1\} = 0.368,$$
$$p_{01} = P\{D_{t+1} = 2\} = 0.184.$$

Una transición de $X_t = 0$ a $X_{t+1} = 0$ implica que la demanda de cámaras en la semana $t + 1$ es 3 o más, después que se agregaron tres cámaras al inventario agotado al principio de la semana, de manera que

$$p_{00} = P\{D_{t+1} \geq 3\} = 0.080.$$

En el caso de los otros dos renglones de **P**, la fórmula al final de la sección 16.1 del siguiente estado es

$$X_{t+1} = \text{máx } \{X_t - D_{t+1}, 0\} \qquad \text{si} \qquad X_t \geq 1.$$

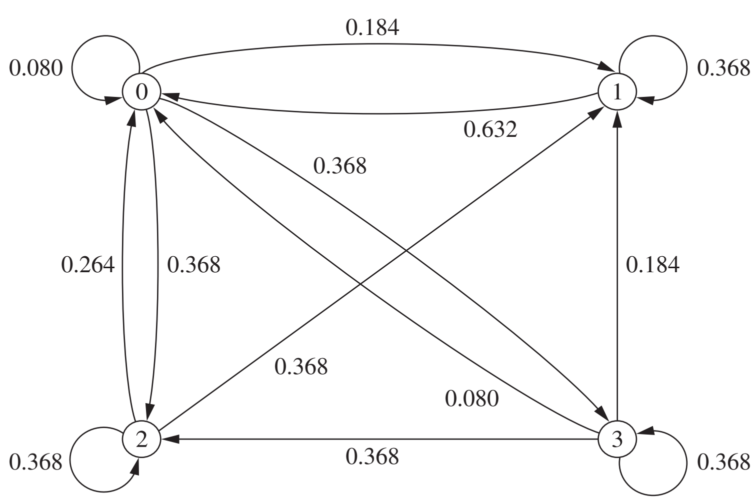
Esto implica que $X_{t+1} \leq X_t$ entonces, $p_{12} = 0$, $p_{13} = 0$ y $p_{23} = 0$. En el caso de las otras transiciones,

$$p_{11} = P\{D_{t+1} = 0\} = 0.368,$$
$$p_{10} = P\{D_{t+1} \geq 1) = 1 - P\{D_{t+1} = 0\} = 0.632,$$
$$p_{22} = P\{D_{t+1} = 0\} = 0.368,$$
$$p_{21} = P\{D_{t+1} = 1\} = 0.368,$$
$$p_{20} = P\{D_{t+1} \geq 2\} = 1 - P\{D_{t+1} \leq 1\} = 1 - (0.368 + 0.368) = 0.264.$$

En el último renglón de **P**, la semana $t + 1$ comienza con tres cámaras en inventario y los cálculos de las probabilidades de transición son exactamente los mismos que las del primer renglón. En consecuencia, la matriz de transición completa (expresándola con tres dígitos significativos) es

$$\mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.080 & 0.184 & 0.368 & 0.368 \\ 0.632 & 0.368 & 0 & 0 \\ 0.264 & 0.368 & 0.368 & 0 \\ 0.080 & 0.184 & 0.368 & 0.368 \end{bmatrix} \end{array}$$

La información que proporciona la matriz de transición también se puede describir mediante el diagrama de transición de estados de la figura 16.2. En el diagrama, los cuatro estados posibles del número de cámaras que se tienen al final de la semana se representan por medio de los cuatro nodos (círculos). Las flechas muestran las transiciones posibles de un estado a otro o, en ocasiones,



■ **FIGURA 16.2**
Diagrama de transición de estados del ejemplo de inventarios.

de regreso al mismo estado, cuando la tienda de cámaras transita del final de una semana al final de la siguiente. El número junto a cada flecha proporciona la probabilidad de que ocurra esa transición en particular cuando la tienda de cámaras tiene el estado que está en la base de la flecha.

Ejemplos adicionales de cadenas de Markov

Ejemplo de acciones. Considere el siguiente modelo del valor de una acción. Al final de un día dado se registra el precio. Si la acción subió, la probabilidad de que suba mañana es de **0.7**. Si la acción bajó, la probabilidad de que suba mañana es de sólo **0.5**. (Para simplificar, cuando la acción permanezca con el mismo precio se considerará un aumento.) Ésta es una cadena de Markov, donde los estados posibles de cada día son los siguientes:

Estado 0: El precio de la acción subió este día.
Estado 1: El precio de la acción bajó este día.

La matriz de transición que muestra cada probabilidad de ir desde un estado particular hoy hasta un estado particular mañana, está dada por

$$\mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{bmatrix} \end{array}$$

La forma del diagrama de transición de estado de este ejemplo es exactamente la misma que la del ejemplo 16.1, por lo que no se repetirá aquí. La única diferencia radica en que las probabilidades de transición en el diagrama son ligeramente diferentes (0.7 reemplaza a 0.8, 0.3 reemplaza a 0.2 y 0.5 reemplaza a 0.6 y 0.4 en la figura 16.1).

Segundo ejemplo de acciones. Suponga ahora que el modelo del mercado de acciones se cambia de manera que el hecho de que una acción suba mañana depende de que haya subido hoy y ayer. En particular, si la acción subió los dos días, ayer *y* hoy, la probabilidad de que suba mañana es de **0.9**. Si la acción subió hoy pero ayer bajó, la probabilidad de que mañana suba es de **0.6**. Si la acción bajó hoy pero ayer subió, la probabilidad de que mañana suba es de **0.5**. Por último, si bajó durante estos dos días, la probabilidad de que mañana suba es de **0.3**. Si se define el estado como la representación del hecho de que la acción baje o suba hoy, el sistema ya no es una cadena de Markov. Sin embargo, se puede transformar en una de ellas si se definen los estados como sigue:[2]

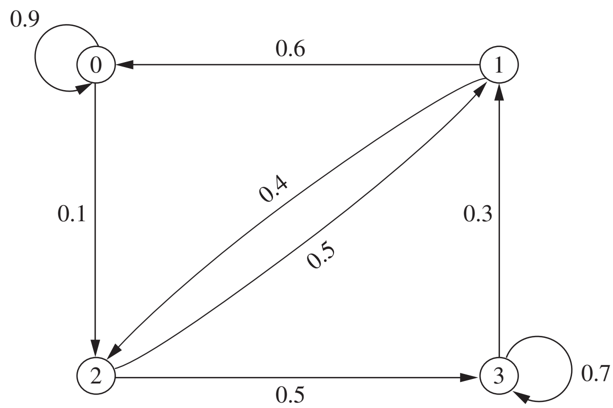
Estado 0: la acción aumentó hoy y ayer.
Estado 1: la acción aumentó hoy y ayer bajó.
Estado 2: la acción bajó hoy y ayer aumentó.
Estado 3: la acción bajó hoy y ayer.

Estos datos conducen a una cadena de Markov de cuatro estados con la siguiente matriz de transición:

$$\mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.9 & 0 & 0.1 & 0 \\ 0.6 & 0 & 0.4 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0.3 & 0 & 0.7 \end{bmatrix} \end{array}$$

La figura 16.3 muestra el diagrama de transición de estado de este ejemplo. Una característica interesante del ejemplo que revela este diagrama y todos los valores de 0 de la matriz de transición es que gran parte de las transiciones del estado i al j son imposibles en un solo paso. En otras palabras, $p_{ij} = 0$ para 8 de las 16 entradas de la matriz de transición. Sin embargo, observe cómo

[2] De nuevo se considera un descenso cuando el nivel de inventarios se mantiene igual. Este ejemplo demuestra que las cadenas de Markov pueden incorporar una cantidad arbitraria de historia, pero a costa de aumentar en forma significativa el número de estados.



■ **FIGURA 16.3**
Diagrama de transición de estado para el segundo ejemplo de inventarios.

siempre es posible ir de cualquier estado i a cualquier estado j (incluyendo $j = i$) en dos pasos. Lo mismo es válido en el caso de tres pasos, cuatro pasos, etc. Por lo tanto, $p_{ij}^{(n)} > 0$ para $n = 2,3, \ldots$ para toda i y j.

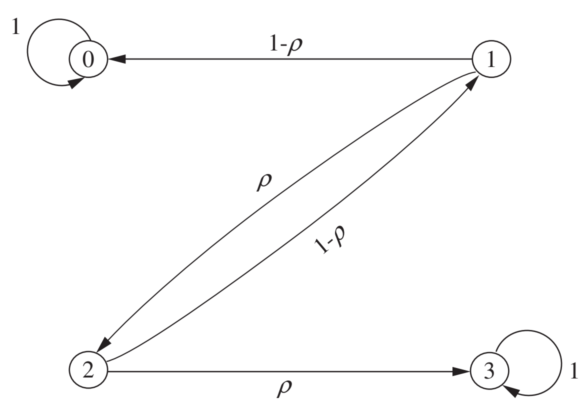
Ejemplo de juego. Otro ejemplo se refiere al juego. Suponga que un jugador tiene 1 dólar y que cada jugada gana 1 dólar con probabilidad $p > 0$ o pierde 1 dólar con probabilidad $1 - p > 0$. El juego termina cuando el jugador acumula 3 dólares o cuando quiebra. Este modelo es una cadena de Markov en la que los estados representan la fortuna del jugador, esto es, 0, 1, 2 o 3 dólares, con la matriz de transición dada por

$$\mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 \\ 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

El diagrama de transición de estado de este ejemplo se muestra en la figura 16.4. Dicho diagrama demuestra que una vez que el proceso entra a un estado 0 o 3, permanecerá en ese estado indefinidamente, puesto que $p_{00} = 1$ y $p_{33} = 1$. Los estados 0 y 3 son ejemplos de lo que se conoce como **estado absorbente** (un estado que nunca se deja una vez que el proceso entra en él). Nos enfocaremos en el análisis de los estados absorbente en la sección 16.7.

Observe que tanto en el ejemplo de inventario como en el del jugador, las etiquetas numéricas de los estados que alcanza el proceso coinciden con la expresión física del sistema, es decir, los niveles de inventario real y la fortuna del jugador, respectivamente, mientras que las etiquetas numéricas de los estados en el ejemplo de la acción no tienen significado físico.

■ **FIGURA 16.4**
Diagrama de transición de estado de un ejemplo de juegos.



16.3 ECUACIONES DE CHAPMAN-KOLMOGOROV

En la sección 16.2 se introdujo la probabilidad de transición de n pasos $p_{ij}^{(n)}$. Las *ecuaciones de Chapman-Kolmogorov* proporcionan un método para calcular estas probabilidades de transición de n pasos:

$$p_{ij}^{(n)} = \sum_{k=0}^{M} p_{ik}^{(m)} p_{kj}^{(n-m)}, \quad \text{para toda } i = 0, 1, \ldots, M, \quad j = 0, 1, \ldots, M, \quad \text{y cualquier } m = 1, 2, \ldots, n-1, \quad n = m+1, m+2, \ldots$$[3]

Estas ecuaciones simplemente señalan que al ir del estado i al estado j en n pasos, el proceso estará en algún estado k después de exactamente m (menor que n) pasos. Así, $p_{ik}^{(m)} p_{kj}^{(n-m)}$ es sólo la probabilidad condicional de que, si comienza en el estado i, el proceso vaya al estado k después de m pasos y después al estado j en $n - m$ pasos. Por lo tanto, al resumir estas probabilidades condicionales sobre todos los estados posibles k se debe obtener $p_{ij}^{(n)}$. Los casos especiales de $m = 1$ y $m = n - 1$ conducen a las expresiones

$$p_{ij}^{(n)} = \sum_{k=0}^{M} p_{ik} p_{kj}^{(n-1)}$$

y

$$p_{ij}^{(n)} = \sum_{k=0}^{M} p_{ik}^{(n-1)} p_{kj},$$

para todos los estados i y j. Estas expresiones permiten que las probabilidades de transición de n pasos se puedan obtener a partir de las probabilidades de transición de un paso de manera recursiva. Esta relación recursiva se explica mejor con la notación matricial (vea el apéndice 4). Para $n = 2$, estas expresiones se convierten en

$$p_{ij}^{(2)} = \sum_{k=0}^{M} p_{ik} p_{kj}, \qquad \text{para todos los estados } i \text{ y } j$$

donde las $p_{ij}^{(2)}$ son los elementos de la matriz $\mathbf{P}^{(2)}$. También note que estos elementos se obtienen al multiplicar la matriz de transición de un paso por sí misma; esto es,

$$\mathbf{P}^{(2)} = \mathbf{P} \cdot \mathbf{P} = \mathbf{P}^2.$$

De la misma manera, las expresiones anteriores de $p_{ij}^{(n)}$ cuando $m = 1$ y $m = n - 1$ indican que la matriz de probabilidades de transición de n pasos es

$$\begin{aligned} \mathbf{P}^{(n)} &= \mathbf{P}\mathbf{P}^{(n-1)} = \mathbf{P}^{(n-1)}\mathbf{P} \\ &= \mathbf{P}\mathbf{P}^{n-1} = \mathbf{P}^{n-1}\mathbf{P} \\ &= \mathbf{P}^n. \end{aligned}$$

Entonces, la matriz de probabilidades de transición de n pasos $\mathbf{P}^n$ se puede obtener al calcular la n-ésima potencia de la matriz de transición de un paso $\mathbf{P}$.

Matrices de transición de *n* pasos del ejemplo del clima

En el caso del ejemplo del clima que se presentó en la sección 16.1, ahora se usarán las fórmulas anteriores para calcular las diferentes matrices de transición de n pasos a partir de la matriz de transición $\mathbf{P}$ (de un paso) que se obtuvo en la sección 16.2. Para iniciar, la matriz de transición de dos pasos es

$$\mathbf{P}^{(2)} = \mathbf{P} \cdot \mathbf{P} = \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} = \begin{bmatrix} 0.76 & 0.24 \\ 0.72 & 0.28 \end{bmatrix}.$$

[3] Estas ecuaciones también se cumplen en el caso trivial, cuando $m = 0$ o $m = n$, pero $m = 1, 2, \ldots, n - 1$ son los únicos casos de interés.

Así, si el clima está en el estado 0 (seco) en un día particular, la probabilidad de estar en el estado 0 dos días después es 0.76, por lo que la probabilidad de estar en el estado 1 (lluvia) es 0.24. En forma similar, si el clima está en el estado 1 ahora, la probabilidad de estar en el estado 0 dos días después es 0.72 mientras que la probabilidad de estar en el estado 1 es 0.28.

Las probabilidades del estado del clima tres, cuatro o cinco días a futuro también se pueden leer de la misma forma a partir de las matrices de transición de tres, cuatro y cinco pasos que se calculan con tres dígitos significativos a continuación.

$$\mathbf{P}^{(3)} = \mathbf{P}^3 = \mathbf{P} \cdot \mathbf{P}^2 = \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} \begin{bmatrix} 0.76 & 0.24 \\ 0.72 & 0.28 \end{bmatrix} = \begin{bmatrix} 0.752 & 0.248 \\ 0.744 & 0.256 \end{bmatrix}$$

$$\mathbf{P}^{(4)} = \mathbf{P}^4 = \mathbf{P} \cdot \mathbf{P}^3 = \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} \begin{bmatrix} 0.752 & 0.248 \\ 0.744 & 0.256 \end{bmatrix} = \begin{bmatrix} 0.75 & 0.25 \\ 0.749 & 0.251 \end{bmatrix}$$

$$\mathbf{P}^{(5)} = \mathbf{P}^5 = \mathbf{P} \cdot \mathbf{P}^4 = \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} \begin{bmatrix} 0.75 & 0.25 \\ 0.749 & 0.251 \end{bmatrix} = \begin{bmatrix} 0.75 & 0.25 \\ 0.75 & 0.25 \end{bmatrix}$$

Observe que la matriz de transición de cinco pasos tiene la interesante característica de que los dos renglones poseen elementos idénticos. Ello refleja el hecho de que la probabilidad del clima que está en un estado particular es en esencia independiente del estado del clima cinco días antes. Por lo tanto, las probabilidades de cualquier renglón de esta matriz de transición de cinco pasos se denominan *probabilidades del estado estable* de la cadena de Markov.

Más adelante se hará un estudio más profundo de las probabilidades del estado estable de una cadena de Markov, donde se incluye cómo obtenerlas de un modo más directo, al inicio de la sección 16.5.

Matrices de transición de *n* pasos del ejemplo de inventarios

De nuevo en el ejemplo de inventarios, incluido en la sección 16.1, calcularemos ahora sus matrices de transición de n pasos con tres dígitos decimales con $n = 2$, 4 y 8. Para comenzar, su matriz de transición de un solo paso **P** que se obtuvo en la sección 16.2 puede usarse para obtener la matriz de transición de dos pasos $\mathbf{P}^{(2)}$ de la siguiente forma:

$$\mathbf{P}^{(2)} = \mathbf{P}^2 = \begin{bmatrix} 0.080 & 0.184 & 0.368 & 0.368 \\ 0.632 & 0.368 & 0 & 0 \\ 0.264 & 0.368 & 0.368 & 0 \\ 0.080 & 0.184 & 0.368 & 0.368 \end{bmatrix} \begin{bmatrix} 0.080 & 0.184 & 0.368 & 0.368 \\ 0.632 & 0.368 & 0 & 0 \\ 0.264 & 0.368 & 0.368 & 0 \\ 0.080 & 0.184 & 0.368 & 0.368 \end{bmatrix}$$

$$= \begin{bmatrix} 0.249 & 0.286 & 0.300 & 0.165 \\ 0.283 & 0.252 & 0.233 & 0.233 \\ 0.351 & 0.319 & 0.233 & 0.097 \\ 0.249 & 0.286 & 0.300 & 0.165 \end{bmatrix}.$$

Por ejemplo, dado que se tiene una cámara en existencia al final de la semana, la probabilidad de que no haya cámaras en inventario dos semanas después es de 0.283, esto es $p_{10}^{(2)} = 0.283$. De manera similar, dado que se tienen dos cámaras al final de una semana, la probabilidad de que haya tres cámaras en inventario dos semanas después es de 0.097; esto es, $p_{23}^{(2)} = 0.097$.

La matriz de transición de cuatro pasos también se puede obtener de la siguiente manera:

$$\mathbf{P}^{(4)} = \mathbf{P}^4 = \mathbf{P}^{(2)} \cdot \mathbf{P}^{(2)}$$

$$= \begin{bmatrix} 0.249 & 0.286 & 0.300 & 0.165 \\ 0.283 & 0.252 & 0.233 & 0.233 \\ 0.351 & 0.319 & 0.233 & 0.097 \\ 0.249 & 0.286 & 0.300 & 0.165 \end{bmatrix} \begin{bmatrix} 0.249 & 0.286 & 0.300 & 0.165 \\ 0.283 & 0.252 & 0.233 & 0.233 \\ 0.351 & 0.319 & 0.233 & 0.097 \\ 0.249 & 0.286 & 0.300 & 0.165 \end{bmatrix}$$

$$= \begin{bmatrix} 0.289 & 0.286 & 0.261 & 0.164 \\ 0.282 & 0.285 & 0.268 & 0.166 \\ 0.284 & 0.283 & 0.263 & 0.171 \\ 0.289 & 0.286 & 0.261 & 0.164 \end{bmatrix}.$$

Por ejemplo, dado que queda una cámara al final de una semana, 0.282 es la probabilidad de que no haya cámaras en inventario cuatro semanas más tarde; es decir, $p_{10}^{(4)} = 0.282$. De igual manera, dado que quedan dos cámaras en el almacén al final de cierta semana, se tiene una probabilidad de 0.171 de que haya tres cámaras en el almacén cuatro semanas después; esto es, $p_{23}^{(4)} = 0.171$.

Las probabilidades de transición del número de cámaras en inventario dentro de ocho semanas se puede leer de la misma forma a partir de la matriz de transición de ocho pasos que se calcula a continuación.

$$\mathbf{P}^{(8)} = \mathbf{P}^8 = \mathbf{P}^{(4)} \cdot \mathbf{P}^{(4)}$$

$$= \begin{bmatrix} 0.289 & 0.286 & 0.261 & 0.164 \\ 0.282 & 0.285 & 0.268 & 0.166 \\ 0.284 & 0.283 & 0.263 & 0.171 \\ 0.289 & 0.286 & 0.261 & 0.164 \end{bmatrix} \begin{bmatrix} 0.289 & 0.286 & 0.261 & 0.164 \\ 0.282 & 0.285 & 0.268 & 0.166 \\ 0.284 & 0.283 & 0.263 & 0.171 \\ 0.289 & 0.286 & 0.261 & 0.164 \end{bmatrix}$$

$$\begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.286 & 0.285 & 0.264 & 0.166 \\ 0.286 & 0.285 & 0.264 & 0.166 \\ 0.286 & 0.285 & 0.264 & 0166 \\ 0.286 & 0.285 & 0.264 & 0.166 \end{bmatrix} \end{array}$$

De igual modo que en la matriz de transición de cinco pasos del ejemplo del clima, esta matriz tiene la interesante característica de que sus renglones tienen elementos idénticos. De nuevo, la razón es que las probabilidades en cualquier renglón son las *probabilidades del estado estable* de esta cadena de Markov, es decir, las probabilidades del estado del sistema después de un tiempo suficiente han llegado a un punto en el que el estado inicial ya no es relevante.

En el IOR Tutorial se incluye una rutina para calcular $\mathbf{P}^{(n)} = \mathbf{P}^n$ para cualquier entero positivo $n \leq 99$.

Probabilidades de estado incondicionales

Recuerde que las probabilidades de transición de uno o de n pasos son probabilidades *condicionales*; por ejemplo, $P\{X_n = j|X_0 = i\} = p_{ij}^{(n)}$. Se supone que n es lo suficientemente pequeña como para que estas probabilidades todavía no sean las del *estado* estable. En este caso, si se desea la probabilidad *incondicional* $P\{X_n = j\}$, es necesario que se especifique la distribución de probabilidad del estado inicial, o sea, $p\{X_0 = i\}$ para $i = 0, 1, \ldots, M$. Entonces

$$P\{X_n = j\} = P\{X_0 = 0\}\, p_{0j}^{(n)} + P\{X_0 = 1\}p_{1j}^{(n)} + \cdots + P\{X_0 = M\}p_{Mj}^{(n)}.$$

En el ejemplo de inventarios se supuso que al inicio se contaba con tres unidades en inventario, es decir, $X_0 = 3$. Así, $P\{X_0 = 0\} = P\{X_0 = 1\} = P\{X_0 = 2\} = 0$ y $P\{X_0 = 3\} = 1$. Por lo tanto, la probabilidad (incondicional) de que haya tres cámaras en inventario dos semanas después de que el sistema se puso en marcha es $P\{X_2 = 3\} = (1)p_{33}^{(2)} = 0.165$.

16.4 CLASIFICACIÓN DE ESTADOS EN UNA CADENA DE MARKOV

Acabamos de ver en la parte final de la sección anterior que las probabilidades de transición de n pasos del ejemplo del inventario convergen hacia las probabilidades del estado estable después de un número de pasos suficiente. Sin embargo, esto no es válido para todas las cadenas de Markov. Las propiedades a largo plazo de una cadena de Markov dependen en gran medida de las características de sus estados y de la matriz de transición. Para describir con más detalle las propiedades de una cadena de Markov es necesario presentar algunos conceptos y definiciones que se refieren a estos estados.

Se dice que el estado j es **accesible** desde el estado i si $p_{ij}^{(n)} > 0$ para alguna $n \geq 0$. (Recuerde que $p_{ij}^{(n)}$ es sólo la probabilidad condicional de llegar al estado j después de n pasos, si el sistema está en el estado i.) Entonces, que el estado j sea accesible desde el estado i significa que es posible que el sistema llegue finalmente al estado j si comienza en el estado i. Esto es particularmente vá-

lido para el ejemplo del clima (vea la figura 16.1) puesto que $p_{ij} > 0$ para toda i y j. En el ejemplo de inventarios (vea la figura 16.2), $p_{ij}^{(2)} > 0$ para todo i y j de manera que cada estado es accesible desde cualquier otro estado. En general, una condición suficiente para que *todos* los estados sean accesibles es que exista un valor de n para el que $p_{ij}^{(n)} > 0$ para todo i y j.

En el ejemplo del juego que se presentó al final de la sección 16.2 (vea la figura 16.4), el estado 2 no es accesible desde el estado 3. Esto se puede deducir del contexto del juego (una vez que el jugador llega al estado 3 nunca lo deja), lo que implica que $p_{32}^{(n)} = 0$ para toda $n \geq 0$. Sin embargo, aun cuando el estado 2 *no* es accesible desde el estado 3, el estado 3 sí *es* accesible desde el estado 2 puesto que, con $n = 1$, la matriz de transición del final de la sección 16.2 indica que $p_{23} = p > 0$.

Si el estado j es accesible desde el estado i y el estado i es accesible desde el estado j, entonces se dice que los estados i y j se **comunican**. En el ejemplo de inventarios, todos los estados se comunican. En el ejemplo del juego, los estados 2 y 3 no se comunican. En general:

1. Cualquier estado se comunica consigo mismo (porque $p_{ii}^{(0)} = P\{X_0 = i|X_0 = i\} = 1$).
2. Si el estado i se comunica con el estado j, entonces el estado j se comunica con el estado i.
3. Si el estado i se comunica con el estado j y éste con el estado k, entonces el estado i se comunica con el estado k.

Las propiedades 1 y 2 se deducen de la definición de estados que se comunican, mientras que la propiedad 3 se deriva de las ecuaciones de Chapman-Kolmogorov.

Como resultado de estas propiedades de comunicación se puede hacer una partición del espacio de estados en **clases** separadas, donde se dice que dos estados que se comunican pertenecen a la misma clase. (Una clase puede consistir en un solo estado.) Si existe sólo una clase, es decir, si todos los estados se comunican, se dice que la cadena de Markov es **irreducible**. Tanto en el ejemplo del clima como en el de inventarios, la cadena de Markov es irreducible. En el primer ejemplo de las acciones (sección 16.2), la cadena de Markov también es irreducible. El ejemplo del juego contiene tres clases; el estado 0 forma una clase, el estado 3 forma otra y los estados 1 y 2 forman una tercera clase.

Estados recurrentes y estados transitorios

Con frecuencia es útil saber si un proceso que comienza en un estado regresará alguna vez a él. La siguiente es una posibilidad.

> Un estado se llama estado **transitorio** si, después de haber entrado a este estado, el proceso *nunca regresa* a él. Por consiguiente, el estado i es transitorio si y sólo si existe un estado j ($j \neq i$) que es accesible desde el estado i, pero no viceversa, esto es, el estado i no es accesible desde el estado j.

Así, si el estado i es transitorio y el proceso visita este estado, existe una probabilidad positiva (quizá incluso de 1) de que el proceso se moverá al estado j y nunca regresará al estado i. En consecuencia, un estado transitorio será visitado sólo un número finito de veces. Para ilustrar lo anterior, considere el ejemplo del juego que se presentó al final de la sección 16.2. Su diagrama de transición de estado que se muestra en la figura 16.4 indica que ambos estados (1 y 2) son transitorios ya que el proceso los abandonará tarde o temprano para entrar al estado 0 o al 3 y, después, permanecerá en dicho estado de manera indefinida.

Cuando se inicia en el estado i, otra posibilidad es que el proceso *definitivamente* regrese a ese estado.

> Se dice que un estado es **recurrente** si, después de haber entrado a este estado, el proceso *definitivamente regresará* a ese estado. Por consiguiente, un estado es recurrente si y sólo si no es transitorio.

Como un estado recurrente será visitado de nuevo después de cada visita, podría ser visitado un número infinito de veces si el proceso continuara por siempre. Por ejemplo, todos los estados en los diagramas de transición de estado que se muestran en las figuras 16.1, 16.2 y 16.3 son estados recurrentes debido a que el proceso siempre regresará a cada uno de ellos. Inclusive en el ejemplo del juego, los estados 0 y 3 son recurrentes debido a que el proceso se mantendrá regresando de manera inmediata a uno de estos estados en forma indefinida, una vez que el proceso haya entrado a ese estado. Observe en la figura 16.4 cómo finalmente el proceso entrará a cualquiera de los estados 0 y 3 y, después, ya nunca los abandonará.

Si el proceso entra en cierto estado y permanece en él al siguiente paso, se considera un *regreso* a ese estado. En consecuencia, el siguiente tipo de estado se considera un tipo especial de estado recurrente.

> Un estado se llama estado **absorbente** si, después de haber entrado ahí, el proceso *nunca saldrá* de él. Por consiguiente, el estado i es un estado absorbente si y sólo si $p_{ii} = 1$.

Como se acaba de mencionar, tanto el estado 0 como el 3 del ejemplo del juego están de acuerdo con esta definición, por lo que ambos son estados absorbentes, así como también un tipo especial de estado recurrente. En la sección 16.7 se estudiarán los estados absorbentes con más detalle.

La recurrencia es una propiedad de clase. Es decir, todos los estados de una clase son recurrentes o son transitorios. Más aún, en una cadena de Markov de estado finito, no todos los estados pueden ser transitorios. Entonces, todos los estados de una cadena de Markov de estado finito irreducible son recurrentes. Sin duda, se puede identificar una cadena de Markov de estado finito irreducible (y, por lo tanto, concluir que todos los estados son recurrentes) demostrando que todos los estados del proceso se comunican. Ya se hizo notar que una condición suficiente para que *todos* los estados sean accesibles (y, por lo tanto, se comuniquen unos con otros) es que exista un valor de n para el cual $p_{ij}^{(n)} > 0$ para toda i y j. En este contexto, todos los estados del ejemplo de inventarios (vea la figura 16.2) son recurrentes, puesto que $p_{ij}^{(2)}$ es positiva para toda i y j. De manera parecida, el primer ejemplo sobre las acciones contiene sólo estados recurrentes, puesto que p_{ij} es positiva para toda i y j. Cuando se calcula $p_{ij}^{(2)}$ para toda i y j en el segundo ejemplo de acciones de la sección 16.2 (vea la figura 16.3) se concluye que todos los estados son recurrentes porque $p_{ij}^{(2)} > 0$ para toda i y j.

Como otro ejemplo, suponga que un proceso de Markov tiene la siguiente matriz de transición:

$$
\mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array}
\begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\
\left[\begin{array}{ccccc}
\frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\
1 & 0 & 0 & 0 & 0
\end{array}\right] \end{array}
$$

Observe que el estado 2 es absorbente (y, por lo tanto, recurrente), porque si el proceso entra en él (tercer renglón de la matriz), nunca sale. El estado 3 es transitorio porque una vez que el proceso se encuentra en él, existe una probabilidad positiva de nunca regresar. La probabilidad de que el proceso vaya del estado 3 al estado 2 en el primer paso es $\frac{1}{3}$. Si el proceso está en el estado 2, permanece en ese estado. Cuando el proceso deja el estado 4, nunca vuelve. Los estados 0 y 1 son recurrentes. Para comprobar todo lo anterior, observe en **P** que si el proceso comienza en cualquier de estos estados, nunca sale de ellos. Aún más, cuando el proceso se mueve de uno de estos estados al otro, siempre regresa al estado original.

Propiedades de periodicidad

Otra propiedad útil de las cadenas de Markov es la de *periodicidad*. El **periodo** de un estado i se define como el entero t ($t > 1$) si $p_{ii}^{(n)} = 0$ para todos los valores de n distintos de $t, 2t, 3t, \ldots$, y t es el entero más grande con esta propiedad. En el problema del juego (fin de la sección 16.2), al comenzar en el estado 1, es posible que el proceso entre al estado 1 sólo en los tiempos 2, 4, . . ., en cuyo caso se dice que el estado 1 tiene periodo 2. Esto es evidente si se observa que el jugador puede salir a mano (ni ganar ni perder) sólo en los tiempos 2, 4, . . ., lo que se puede verificar si se calcula $p_{11}^{(n)}$ para toda n y se observa que $p_{11}^{(n)} = 0$ para n impar. También es posible ver en la figura 16.4 que el proceso siempre toma dos pasos para regresar al estado 1 hasta que el proceso es absorbido en cualquiera de los estados 0 o 3. (La misma conclusión también se aplica al estado 2.)

Si existen dos números consecutivos s y $s + 1$ tales que el proceso puede encontrarse en el estado i en los tiempos s y $s + 1$, se dice que el estado tiene periodo 1 y se llama **aperiódico**.

Igual que la recurrencia es una propiedad de clase, se puede demostrar que la periodicidad también lo es. Esto es, si el estado i de una clase tiene periodo t, todos los estados de esa clase tienen

periodo t. En el ejemplo del juego, el estado 2 tiene periodo 2 porque está en la misma clase que el estado 1 y, como se vio, el estado 1 tiene periodo 2.

Es posible que una cadena de Markov tenga tanto una clase de estados recurrentes como una de estados transitorios donde las dos clases tienen diferentes periodos mayores que 1. Si el lector desea ver una cadena de Markov donde se presente esta situación, **otro ejemplo** de este tipo se proporciona en la sección Worked Examples del sitio en internet de este libro.

En una cadena de Markov de estado finito, los estados recurrentes aperiódicos se llaman **ergódicos**. Se dice que una cadena de Markov es *ergódica* si todos sus estados son ergódicos. En seguida se verá que una propiedad clave a largo plazo de una cadena de Markov que es tanto irreducible como ergódica es que sus probabilidades de transición de n pasos convergirán a las probabilidades del estado estable conforme n se haga más grande.

16.5 PROPIEDADES A LARGO PLAZO DE LAS CADENAS DE MARKOV

Probabilidades de estado estable

Mientras se calculaban las matrices de transición de n pasos de los ejemplos del clima y de inventarios en la sección 16.3, se observó una característica interesante de estas matrices. Si n es lo suficientemente grande ($n = 5$ en el ejemplo del clima y $n = 8$ en el ejemplo de inventarios), todos los renglones de la matriz tienen elementos idénticos, lo que significa que la probabilidad de que el sistema esté en cada estado j ya no depende del estado inicial del sistema. En otras palabras, parece que existe una probabilidad límite de que el sistema se encuentre en el estado j después de un número grande de transiciones, y que esta probabilidad es independiente del estado inicial. En realidad, estas propiedades del comportamiento a largo plazo de un proceso de Markov de estado finito se cumplen en condiciones relativamente generales, como se resume a continuación.

Para una cadena de Markov irreducible ergódica el $\lim\limits_{n\to\infty} p_{ij}^{(n)}$ existe y es independiente de i.

Aún más,

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j > 0,$$

donde las π_j satisfacen de manera única las siguientes **ecuaciones de estado estable**

$$\pi_j = \sum_{i=0}^{M} \pi_i p_{ij}, \qquad \text{para } j = 0, 1, \ldots, M,$$

$$\sum_{j=0}^{M} \pi_j = 1.$$

Si usted prefiere trabajar con un sistema de ecuaciones en forma matricial, este sistema (excluyendo la ecuación sum = 1) también puede expresarse como

$$\pi_j = \pi \mathbf{P}.$$

donde $\pi = (\pi_0, \pi_1, \ldots, \pi_M)$.

Las π_j se llaman **probabilidades de estado estable** de la cadena de Markov. El término *probabilidad de estado estable* significa que la probabilidad de encontrar el proceso en cierto estado, por ejemplo j, después de un número grande de transiciones tiende al valor π_j, y es independiente de la distribución de probabilidad inicial definida para los estados. Es importante observar que la probabilidad de estado estable *no* significa que el proceso se establezca en un estado. Por el contrario, el proceso continúa haciendo transiciones de un estado a otro y en cualquier paso n la probabilidad de transición del estado i al estado j es todavía p_{ij}.

También se puede interpretar las π_j como *probabilidades estacionarias* (que no deben confundirse con las probabilidades de transición estacionarias) en el siguiente sentido. Si la probabilidad *inicial* de encontrarse en estado j está dada por π_j (esto es, $P\{X_0 = j\} = \pi_j$) para toda j, entonces la probabilidad de encontrar el proceso en el estado j en el tiempo $n = 1, 2, \ldots$ también está dada por π_j (es decir, $P\{X_n = j\} = \pi_j$).

Debe observarse que las ecuaciones de estado estable consisten en $M + 2$ ecuaciones con $M + 1$ incógnitas. Como el sistema tiene una solución única, al menos una de las ecuaciones debe ser redundante, por lo que se puede eliminar. No puede ser la ecuación

$$\sum_{j=0}^{M} \pi_j = 1,$$

como $\pi_j = 0$ para toda j satisfaría las otras $M + 1$ ecuaciones. Aún más, las soluciones de las otras $M + 1$ ecuaciones de estado estable tienen una solución única con una constante multiplicativa y es la ecuación final la que fuerza la solución a ser una distribución de probabilidad.

Aplicación al ejemplo del clima. El ejemplo del clima que se introdujo en la sección 16.1 y se formuló en la sección 16.2 tiene sólo dos estados (seco y lluvioso), por lo que las ecuaciones anteriores de estado estable se convierten en

$$\begin{aligned}
\pi_0 &= \pi_0 p_{00} + \pi_1 p_{10},\\
\pi_1 &= \pi_0 p_{01} + \pi_1 p_{11},\\
1 &= \pi_0 \quad\; + \pi_1.
\end{aligned}$$

Lo que se intuye detrás de la primera ecuación es que, en el estado estable, la probabilidad de quedar en el estado 0 después de la siguiente transición debe ser igual a 1) la probabilidad de estar en el estado 0 ahora *y* luego permanecer en el estado 0 después de la siguiente transición *más* 2) la probabilidad de estar en el estado 1 ahora *y* luego hacer la transición al estado 0. La lógica de la segunda ecuación es la misma, sólo que está en términos del estado 1. La tercera ecuación sólo expresa el hecho de que las probabilidades de estos estados mutuamente excluyentes deben sumar 1.

En referencia a las probabilidades de transición de la sección 16.2 de este ejemplo, estas ecuaciones se convierten en

$$\begin{aligned}
\pi_0 &= 0.8\pi_0 + 0.6\pi_1, &&\text{así} \qquad 0.2\pi_0 = 0.6\pi_1,\\
\pi_1 &= 0.2\pi_0 + 0.4\pi_1, &&\text{así} \qquad 0.6\pi_1 = 0.2\pi_0,\\
1 &= \pi_0 \quad\; + \pi_1.
\end{aligned}$$

Observe que una de las dos primeras ecuaciones es redundante puesto que ambas ecuaciones se reducen a $\pi_0 = 3\pi_1$. Al combinar estos resultados con la tercera ecuación se producen de inmediato las siguientes probabilidades de estado estable:

$$\pi_0 = 0.25, \qquad \pi_1 = 0.75$$

Éstas son las mismas probabilidades que las que se obtuvieron en cada renglón de la matriz de cinco pasos que se calculó en la sección 16.3, debido a que cinco transiciones probaron ser suficientes para que las probabilidades de estado sean en esencia independientes del estado inicial.

Aplicación al ejemplo de inventarios. El ejemplo de inventarios que se introdujo en la sección 16.1 y se formuló en la sección 16.2 tiene cuatro estados. Por lo tanto, las ecuaciones de estado estable se pueden expresar como

$$\begin{aligned}
\pi_0 &= \pi_0 p_{00} + \pi_1 p_{10} + \pi_2 p_{20} + \pi_3 p_{30},\\
\pi_1 &= \pi_0 p_{01} + \pi_1 p_{11} + \pi_2 p_{21} + \pi_3 p_{31},\\
\pi_2 &= \pi_0 p_{02} + \pi_1 p_{12} + \pi_2 p_{22} + \pi_3 p_{32},\\
\pi_3 &= \pi_0 p_{03} + \pi_1 p_{13} + \pi_2 p_{23} + \pi_3 p_{33},\\
1 &= \pi_0 \quad\; + \pi_1 \quad\; + \pi_2 \quad\; + \pi_3.
\end{aligned}$$

Al sustituir los valores de p_{ij} (vea la matriz de transición en la sección 16.2) en estas ecuaciones se obtiene

$$\begin{aligned}
\pi_0 &= 0.080\pi_0 + 0.632\pi_1 + 0.264\pi_2 + 0.080\pi_3,\\
\pi_1 &= 0.184\pi_0 + 0.368\pi_1 + 0.368\pi_2 + 0.184\pi_3,\\
\pi_2 &= 0.368\pi_0 \qquad\qquad\quad + 0.368\pi_2 + 0.368\pi_3,\\
\pi_3 &= 0.368\pi_0 \qquad\qquad\qquad\qquad\qquad + 0.368\pi_3,\\
1 &= \quad\;\; \pi_0 + \quad\;\; \pi_1 + \quad\;\; \pi_2 + \quad\;\; \pi_3.
\end{aligned}$$

Cuando se resuelven en forma simultánea las últimas cuatro ecuaciones se obtiene la solución

$$\pi_0 = 0.286, \qquad \pi_1 = 0.285, \qquad \pi_2 = 0.263, \qquad \pi_3 = 0.166,$$

que en esencia son los resultados que aparecen en la matriz $\mathbf{P}^{(8)}$ de la sección 16.3. En consecuencia, después de muchas semanas, la probabilidad de encontrar cero, una, dos y tres cámaras en el almacén tiende a 0.286, 0.285, 0.263 y 0.166, respectivamente.

Más acerca de las probabilidades de estado estable. En el IOR Tutorial se incluye una rutina para resolver las ecuaciones de estado estable para obtener las probabilidades de estado estable. Además, en la sección Worked Examples de nuestro sitio en internet se incluye **otro ejemplo** de aplicación de las probabilidades de estado estable (que explica la utilización de la técnica que se describe en la siguiente subsección) para determinar la mejor alternativa entre varias con base en el costo.

Existen otros resultados importantes respecto de las probabilidades de estado estable. En particular, si i y j son estados recurrentes que pertenecen a clases distintas, entonces

$$p_{ij}^{(n)} = 0, \qquad \text{para toda } n.$$

Este resultado es una consecuencia de la definición de clase.

De manera similar, si j es un estado transitorio, entonces

$$\lim_{n\to\infty} p_{ij}^{(n)} = 0, \qquad \text{para toda } i.$$

De este modo, la probabilidad de encontrar el proceso en un estado transitorio después de un número grande de transiciones tiende a cero.

Costo promedio esperado por unidad de tiempo

La subsección anterior estudió las cadenas de Markov de estado finito irreducible cuyos estados son ergódicos (recurrentes y aperiódicos). Si se relaja el requerimiento de que los estados sean aperiódicos, entonces el límite

$$\lim_{n\to\infty} p_{ij}^{(n)}$$

puede no existir. Para ilustrar este punto considere la matriz de transición de dos estados

$$\mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{array}.$$

Si el proceso comienza en el estado 0 en el tiempo 0, estará en el estado 0 en los tiempos 2, 4, 6, . . . y en el estado 1 en los tiempos 1, 3, 5, ... Entonces, $p_{00}^{(n)} = 1$ si n es par y $p_{00}^{(n)} = 1$ si n es impar, de manera que el

$$\lim_{n\to\infty} p_{00}^{(n)}$$

no existe. Sin embargo, el siguiente límite siempre existe para una cadena de Markov irreducible de estado finito:

$$\lim_{n\to\infty} \left(\frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)} \right) = \pi_j,$$

donde las π_j satisfacen las ecuaciones de estado estable de la subsección anterior.

Este resultado es en extremo importante para calcular el *costo promedio a largo plazo por unidad de tiempo* asociado a una cadena de Markov. Suponga que se incurre en un costo (u otra función de penalización) $C(X_t)$ cuando el proceso se encuentra en el estado X_t en el tiempo t, para $t = 0, 1, 2, \ldots$ Observe que $C(X_t)$ es una variable aleatoria que toma cualquiera de los valores $C(0)$,

$C(1), \ldots, C(M)$ y que la función $C(\cdot)$ es independiente de t. El costo promedio esperado en el que se incurre a lo largo de los primeros n periodos está dado por la expresión

$$E\left[\frac{1}{n}\sum_{t=1}^{n} C(X_t)\right].$$

Si se usa el resultado de que

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=1}^{n} p_{ij}^{(k)}\right) = \pi_j,$$

se puede demostrar que el *costo promedio esperado por unidad de tiempo* (a largo plazo), está dado por

$$\lim_{n\to\infty} E\left[\frac{1}{n}\sum_{t=1}^{n} C(X_t)\right] = \sum_{j=0}^{M} \pi_j C(j).$$

Aplicación al ejemplo de inventarios. A manera de ejemplo, considere el problema de inventarios que se presentó en la sección 16.1, donde la solución de las π_j se obtuvo en una subsección anterior. Suponga que la tienda de cámaras encuentra que se debe asignar un cargo por almacenamiento por cada cámara que permanece en la tienda al final de la semana. El costo se carga de la siguiente manera:

$$C(x_t) = \begin{cases} 0 & \text{si} \quad x_t = 0 \\ 2 & \text{si} \quad x_t = 1 \\ 8 & \text{si} \quad x_t = 2 \\ 18 & \text{si} \quad x_t = 3 \end{cases}$$

Mediante el uso de las probabilidades de estado estable que se dedujeron en esta sección, el costo promedio esperado por semana, a largo plazo, por mantener el inventario, se puede obtener de la ecuación anterior; esto es,

$$\lim_{n\to\infty} E\left[\frac{1}{n}\sum_{t=1}^{n} C(X_t)\right] = 0.286(0) + 0.285(2) + 0.263(8) + 0.166(18) = 5.662.$$

Observe que otra medida del costo promedio esperado (a largo plazo) por unidad de tiempo es el *costo promedio real por unidad de tiempo* (a largo plazo). Se puede demostrar que esta última medida está dada por

$$\lim_{n\to\infty} \left[\frac{1}{n}\sum_{t=1}^{n} C(X_t)\right] = \sum_{j=0}^{M} \pi_j C(j)$$

en esencia para todas las trayectorias del proceso. Así, ambas medidas conducen al mismo resultado. Estos resultados también se pueden usar para interpretar el significado de las π_j. Para hacerlo, sea

$$C(X_t) = \begin{cases} 1 & \text{si } X_t = j \\ 0 & \text{si } X_t \neq j. \end{cases}$$

La fracción esperada del número de veces (a largo plazo) que el sistema se encuentra en el estado j está dada entonces por

$$\lim_{n\to\infty} E\left[\frac{1}{n}\sum_{t=1}^{n} C(X_t)\right] = \lim_{n\to\infty} E(\text{fracción de veces que el sistema está en el estado } j) = \pi_j.$$

De igual manera, π_j se puede interpretar también como la fracción o porcentaje real (a largo plazo) del número de veces que el sistema se encuentra en el estado j.

Costo promedio esperado por unidad de tiempo de funciones de costo complejas

En la subsección anterior, la función de costo se basó nada más en el estado en el que se encuentra el proceso en el tiempo t. En muchos problemas importantes el costo también puede depender de otra variable aleatoria.

Por ejemplo, en el problema de inventarios de la sección 16.1, suponga que debe tomarse en cuenta el costo de ordenar y el costo de penalización por demanda insatisfecha (los costos de almacenaje son pequeños, por lo que se pasarán por alto). Es razonable suponer que el número de cámaras ordenadas al principio de la semana t depende sólo del estado del proceso X_{t-1} (el número de cámaras que se tiene) cuando se hace el pedido al final de la semana $t-1$. Sin embargo, el costo de la demanda que no se satisfizo durante la semana t dependerá de la demanda D_t. Por lo tanto, el costo total (costo de ordenar más costo de la demanda insatisfecha) de la semana t es una función de X_{t-1} y de D_t, esto es, $C(X_{t-1}, D_t)$.

Bajo los supuestos de este ejemplo se puede demostrar que (a largo plazo) el *costo promedio esperado por unidad de tiempo* está dado por

$$\lim_{n\to\infty} E\left[\frac{1}{n}\sum_{t=1}^{n} C(X_{t-1}, D_t)\right] = \sum_{j=0}^{M} k(j)\,\pi_j,$$

donde

$$k(j) = E[C(j, D_t)],$$

y esta última esperanza (condicional) se toma respecto de la distribución de probabilidad de la variable aleatoria D_t, dado el estado j. De manera similar, el costo promedio real (a largo plazo) por unidad de tiempo está dado por

$$\lim_{n\to\infty} \left[\frac{1}{n}\sum_{t=1}^{n} C(X_{t-1}, D_t)\right] = \sum_{j=0}^{M} k(j)\pi_j.$$

Se deben asignar valores numéricos a las dos componentes de $C(X_{t-1}, D_t)$ en este ejemplo, es decir, el costo del pedido y el costo de penalización por la demanda insatisfecha. Si se ordena $z > 0$ cámaras, se incurre en un costo de $(10 + 25z)$ dólares. Si no se ordenan cámaras, no hay cargos por ordenar. Cada unidad de demanda insatisfecha (ventas perdidas), tiene un costo de 50 dólares. Entonces, dada la política de ordenar descrita en la sección 16.1, el costo en la semana t está dado por

$$C(X_{t-1}, D_t) = \begin{cases} 10 + (25)(3) + 50\,\text{máx}\{D_t - 3, 0\} & \text{si} \quad X_{t-1} = 0 \\ 50\,\text{máx}\,\{D_t - X_{t-1}, 0\} & \text{si} \quad X_{t-1} \geq 1, \end{cases}$$

para $t = 1, 2, \ldots$ En consecuencia,

$$C(0, D_t) = 85 + 50\,\text{máx}\{D_t - 3, 0\},$$

de manera que

$$\begin{aligned} k(0) = E[C(0, D_t)] &= 85 + 50E(\text{máx}\{D_t - 3, 0\}) \\ &= 85 + 50[P_D(4) + 2P_D(5) + 3P_D(6) + \cdots], \end{aligned}$$

donde $P_D(i)$ es la probabilidad de que la demanda sea igual a i, según una distribución de Poisson con media de 1, de manera que $P_D(i)$ se vuelve despreciable para i mayor que los valores cercanos a 6. Como $P_D(4) = 0.015$, $P_D(5) = 0.003$ y $P_D(6) = 0.001$, se tiene que $k(0) = 86.2$. También, si se usa $P_D(2) = 0.184$ y $P_D(3) = 0.061$ y se calcula en forma similar, se obtienen los siguientes resultados

$$\begin{aligned} k(1) = E[C(1, D_t)] &= 50E(\text{máx}\{D_t - 1, 0\}) \\ &= 50[P_D(2) + 2P_D(3) + 3P_D(4) + \cdots] \\ &= 18.4, \end{aligned}$$

$$\begin{aligned} k(2) = E[C(2, D_t)] &= 50E(\text{máx}\{D_t - 2, 0\}) \\ &= 50[P_D(3) + 2P_D(4) + 3P_D(5) + \cdots] \\ &= 5.2, \end{aligned}$$

y

$$\begin{aligned} k(3) = E[C(3, D_t)] &= 50E(\text{máx}\{D_t - 3, 0\}) \\ &= 50[P_D(4) + 2P_D(5) + 3P_D(6) + \cdots] \\ &= 1.2. \end{aligned}$$

Así, el costo promedio esperado (a largo plazo) por semana está dado por

$$\sum_{j=0}^{3} k(j)\pi_j = 86.2(0.286) + 18.4(0.285) + 5.2(0.263) + 1.2(0.166) = \$31.46.$$

Éste es el costo asociado con la política de ordenar descrita en la sección 16.1. El costo de otras políticas se puede evaluar de manera similar a fin de identificar la política que minimiza el costo promedio esperado semanal.

Los resultados de esta subsección se presentaron sólo en términos del ejemplo de inventarios, pero los resultados (no numéricos) se cumplen para otros problemas siempre y cuando se satisfagan las siguientes condiciones:

1. $\{X_t\}$ es una cadena de Markov irreducible (estado finito).
2. Asociada con esta cadena de Markov se tiene una secuencia de variables aleatorias $\{D_t\}$, independientes e idénticamente distribuidas.
3. Para una m fija, $m = 0, \pm 1, \pm 2, \ldots$, se incurre en un costo $C(X_t, D_{t+m})$ en el tiempo t, para $t = 0, 1, 2. \ldots$
4. La secuencia $X_0, X_1, X_2, \ldots, X_t$ debe ser independiente de D_{t+m}.

En particular, si se satisfacen estas condiciones, entonces

$$\lim_{n\to\infty} E\left[\frac{1}{n}\sum_{t=1}^{n} C(X_t, D_{t+m})\right] = \sum_{j=0}^{M} k(j)\pi_j,$$

donde

$$k(j) = E[C(j, D_{t+m})],$$

y esta última esperanza condicional se toma respecto de la distribución de probabilidad de la variable aleatoria D_t dado el estado j. Más aún,

$$\lim_{n\to\infty} \left[\frac{1}{n}\sum_{t=1}^{n} C(X_t, D_{t+m})\right] = \sum_{j=0}^{M} k(j)\pi_j$$

para las partes esenciales de todas las trayectorias del proceso.

16.6 TIEMPOS DE PRIMERA PASADA

La sección 16.3 se dedicó a encontrar las probabilidades de transición de n pasos del estado i al estado j. Con frecuencia es conveniente poder hacer afirmaciones en términos de probabilidades sobre el número de transiciones que hace el proceso al ir del estado i al *estado j por primera vez*. Este lapso se llama **tiempo de primera pasada** al ir del estado i al estado j. Cuando $j = i$, este tiempo de primera pasada es igual al número de transiciones hasta que el proceso regresa al estado inicial i. En este caso, el tiempo de primera pasada se llama **tiempo de recurrencia** del estado i.

Para ilustrar estas definiciones, reconsidere el ejemplo de inventarios que se presentó en la sección 16.1, donde X_t es el número de cámaras en inventario al final de la semana t y se comienza con $X_0 = 3$. Suponga que ocurrió lo siguiente

$$X_0 = 3, \quad X_1 = 2, \quad X_2 = 1, \quad X_3 = 0, \quad X_4 = 3, \quad X_5 = 1.$$

En este caso, el tiempo de primera pasada para ir del estado 3 al estado 1 es de dos semanas, el tiempo de primera pasada para ir del estado 3 al estado 0 es de tres semanas y el tiempo de recurrencia del estado 3 es de cuatro semanas.

En general, los tiempos de primera pasada son variables aleatorias. Las distribuciones de probabilidad dependen de las probabilidades de transición del proceso. En particular, $f_{ij}^{(n)}$ denota la probabilidad de que el tiempo de primera pasada del estado i al j sea igual a n. Para $n > 1$, este tiempo de primera pasada es n si la primera transición es del estado i a algún estado k ($k \neq j$) y después el tiempo de primera pasada del estado k al estado j es $n - 1$. Por lo tanto, estas probabilidades satisfacen las siguientes relaciones recursivas:

$$f_{ij}^{(1)} = p_{ij}^{(1)} = p_{ij},$$

$$f_{ij}^{(2)} = \sum_{k \neq j} p_{ik} f_{kj}^{(1)},$$

$$f_{ij}^{(n)} = \sum_{k \neq j} p_{ik} f_{kj}^{(n-1)}.$$

Entonces, la probabilidad de un tiempo de primera pasada del estado i al j en n pasos, se puede calcular de manera recursiva a partir de las probabilidades de transición de un paso.

En el ejemplo de inventarios, la distribución de probabilidad de los tiempos de primera pasada al ir del estado 3 al estado 0 se obtiene de las relaciones recursivas como sigue:

$$\begin{aligned} f_{30}^{(1)} &= p_{30} = 0.080, \\ f_{30}^{(2)} &= p_{31} f_{10}^{(1)} + p_{32} f_{20}^{(1)} + p_{33} f_{30}^{(1)} \\ &= 0.184(0.632) + 0.368(0.264) + 0.368(0.080) = 0.243, \\ &\vdots \end{aligned}$$

donde p_{3k} y $f_{k0}^{(1)} = p_{k0}$ se obtienen de la matriz de transición (de un paso) dada en la sección 16.2.

Para i y j fijas, las $f_{ij}^{(n)}$ son números no negativos tales que

$$\sum_{n=1}^{\infty} f_{ij}^{(n)} \leq 1.$$

Sin embargo, esta suma puede ser estrictamente menor que 1, lo que significa que un proceso que al iniciar se encuentra en el estado i puede no alcanzar nunca el estado j. Cuando la suma sí es igual a 1, las $f_{ij}^{(n)}$ (para $n = 1, 2, \ldots$) pueden considerarse como una distribución de probabilidad de la variable aleatoria, el tiempo de primera pasada.

Aunque puede ser tedioso calcular $f_{ij}^{(n)}$ para toda n, es relativamente sencillo obtener el tiempo esperado de primera pasada del estado i al estado j. Sea μ_{ij} esta esperanza, que se define como

$$\mu_{ij} = \begin{cases} \infty & \text{si } \sum_{n=1}^{\infty} f_{ij}^{(n)} < 1 \\ \sum_{n=1}^{\infty} n f_{ij}^{(n)} & \text{si } \sum_{n=1}^{\infty} f_{ij}^{(n)} = 1. \end{cases}$$

Siempre que

$$\sum_{n=1}^{\infty} f_{ij}^{(n)} = 1,$$

μ_{ij} satisface, de manera única, la ecuación

$$\mu_{ij} = 1 + \sum_{k \neq j} p_{ik} \mu_{kj}.$$

Esta ecuación reconoce que la primera transición desde el estado i puede ser al estado j o a algún otro estado k. Si es al estado j, el tiempo de primera pasada es 1. Dado que la primera transición es a algún estado k ($k \neq j$), lo que ocurre con probabilidad p_{ik}, el tiempo esperado de primera pasada condicional del estado i al estado j es $1 + \mu_{kj}$. Cuando se combinan estos hechos, y se suman todas las posibilidades de la primera transición, se llega a esta ecuación.

En el ejemplo del inventario, estas ecuaciones de μ_{ij} se pueden usar para calcular el tiempo esperado hasta que ya no se tengan cámaras en el almacén, dado que el proceso se inicia cuando se

tienen tres cámaras. Este tiempo esperado es igual que el tiempo esperado de primera pasada μ_{30}. Como todos los estados son recurrentes, el sistema de ecuaciones conduce a las expresiones

$$\mu_{30} = 1 + p_{31}\mu_{10} + p_{32}\mu_{20} + p_{33}\mu_{30},$$
$$\mu_{20} = 1 + p_{21}\mu_{10} + p_{22}\mu_{20} + p_{23}\mu_{30},$$
$$\mu_{10} = 1 + p_{11}\mu_{10} + p_{12}\mu_{20} + p_{13}\mu_{30},$$

o

$$\mu_{30} = 1 + 0.184\mu_{10} + 0.368\mu_{20} + 0.368\mu_{30},$$
$$\mu_{20} = 1 + 0.368\mu_{10} + 0.368\mu_{20},$$
$$\mu_{10} = 1 + 0.368\mu_{10}.$$

La solución simultánea de este sistema es

$$\mu_{10} = 1.58 \text{ semanas},$$
$$\mu_{20} = 2.51 \text{ semanas},$$
$$\mu_{30} = 3.50 \text{ semanas},$$

de manera que el tiempo esperado hasta que la tienda se quede sin cámaras es 3.50 semanas. Al hacer estos cálculos para μ_{30} también se obtienen μ_{20} y μ_{10}.

En el caso de μ_{ij} con $j = i$, μ_{ii} es el número esperado de transiciones hasta que el proceso regresa al estado inicial i, y se llama **tiempo esperado de recurrencia** del estado i. Después de obtener las probabilidades de estado estable ($\pi_0, \pi_1, \ldots, \pi_M$) como se describió en la sección anterior, los tiempos esperados de recurrencia se calculan de inmediato como

$$\mu_{ii} = \frac{1}{\pi_i}, \qquad \text{para } i = 0, 1, \ldots, M.$$

Entonces, en el ejemplo de inventario, donde, $\pi_0 = 0.286$, $\pi_1 = 0.285$, $\pi_2 = 0.263$ y $\pi_3 = 0.166$, los tiempos de recurrencia esperados correspondientes son

$$\mu_{00} = \frac{1}{\pi_0} = 3.50 \text{ semanas}, \qquad \mu_{22} = \frac{1}{\pi_2} = 3.80 \text{ semanas},$$
$$\mu_{11} = \frac{1}{\pi_1} = 3.51 \text{ semanas}, \qquad \mu_{33} = \frac{1}{\pi_3} = 6.02 \text{ semanas}.$$

16.7 ESTADOS ABSORBENTES

En la sección 16.4 se señaló que el estado k se llama *estado absorbente* si $p_{kk} = 1$, de manera que una vez que la cadena llega al estado k permanece ahí para siempre. Si k es un estado absorbente y el proceso comienza en el estado i, la probabilidad de llegar en *algún* momento a k se llama **probabilidad de absorción** al estado k, dado que el sistema comenzó en el estado i. Esta probabilidad se denota por f_{ik}.

Si existen dos o más estados absorbentes en una cadena de Markov y es evidente que el proceso será absorbido en uno de estos estados, es deseable encontrar estas probabilidades de absorción. Dichas probabilidades pueden obtenerse con sólo resolver un sistema de ecuaciones lineales que considera todas las posibilidades de la primera transición y después, dada la primera transición, considera la probabilidad condicional de absorción al estado k. En particular, si el estado k es un estado absorbente, el conjunto de probabilidades de absorción f_{ik} satisface el sistema de ecuaciones

$$f_{ik} = \sum_{j=0}^{M} p_{ij} f_{jk}, \qquad \text{para } i = 0, 1, \ldots, M,$$

sujeta a las condiciones

$$f_{kk} = 1,$$
$$f_{ik} = 0, \qquad \text{si el estado } i \text{ es recurrente e } i \neq k.$$

Las probabilidades de absorción son importantes en las caminatas aleatorias. Una **caminata aleatoria** es una cadena de Markov con la propiedad de que, si el sistema se encuentra en el estado i, entonces en una sola transición, o bien permanecerá en i o se moverá a uno de los dos estados inmediatamente adyacentes a i. Por ejemplo, la caminata aleatoria con frecuencia se usa como modelo para situaciones que incluyen juegos de azar.

Un segundo ejemplo de juegos. Para ilustrar el uso de probabilidades de absorción en una caminata aleatoria, considere un ejemplo sobre juegos de azar similar al que se presentó en la sección 16.2, pero ahora suponga que dos jugadores (A y B), con 2 dólares cada uno, aceptan seguir jugando y apostar 1 dólar cada vez hasta que uno de ellos quiebre. La probabilidad de que A gane una apuesta es $\frac{1}{3}$, por lo que la probabilidad de que gane B es $\frac{2}{3}$. El número de dólares que tiene el jugador A antes de cada apuesta (0, 1, 2, 3 o 4) proporciona los estados de una cadena de Markov con la matriz de transición

$$\mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}.$$

Si se inicia en el estado 2, la probabilidad de absorción al estado 0 (A pierde todo su dinero) se puede obtener al resolver para f_{20} a partir del sistema de ecuaciones que se proporcionó al comienzo de la sección,

$$\begin{aligned}
f_{00} &= 1 \text{ (puesto que el estado 0 es un estado absorbente)}, \\
f_{10} &= \frac{2}{3} f_{00} + \frac{1}{3} f_{20}, \\
f_{20} &= \frac{2}{3} f_{10} + \frac{1}{3} f_{30}, \\
f_{30} &= \frac{2}{3} f_{20} + \frac{1}{3} f_{40}, \\
f_{40} &= 0 \text{ (puesto que el estado 4 es un estado absorbente)}.
\end{aligned}$$

De este sistema de ecuaciones se obtiene

$$f_{20} = \frac{2}{3}\left(\frac{2}{3} + \frac{1}{3} f_{20}\right) + \frac{1}{3}\left(\frac{2}{3} f_{20}\right) = \frac{4}{9} + \frac{4}{9} f_{20},$$

que se reduce a $f_{20} = \frac{4}{5}$ como la probabilidad de absorción en el estado 0.

De manera similar, la probabilidad de que A termine con 4 dólares (B quiebre) cuando comienza con 2 dólares (estado 2) se obtiene al obtener f_{24} del sistema de ecuaciones,

$$\begin{aligned}
f_{04} &= 0 \text{ (puesto que 0 es un estado absorbente)}, \\
f_{14} &= \frac{2}{3} f_{04} + \frac{1}{3} f_{24}, \\
f_{24} &= \frac{2}{3} f_{14} + \frac{1}{3} f_{34}, \\
f_{34} &= \frac{2}{3} f_{24} + \frac{1}{3} f_{44}, \\
f_{44} &= 1 \text{ (puesto que 0 es un estado absorbente)}.
\end{aligned}$$

De aquí se obtiene

$$f_{24} = \frac{2}{3}\left(\frac{1}{3} f_{24}\right) + \frac{1}{3}\left(\frac{2}{3} f_{24} + \frac{1}{3}\right) = \frac{4}{9} f_{24} + \frac{1}{9},$$

de manera que $f_{24} = \frac{1}{5}$ es la probabilidad de absorción en el estado 4.

Ejemplo de evaluación de crédito. Existen muchas otras situaciones en las cuales los estados absorbentes tienen un papel importante. Considere una tienda departamental que clasifica el saldo de la cuenta de un cliente como pagada (estado 0), 1 a 30 días de retraso (estado 1), 31 a 60 días de retraso (estado 2) o mala deuda (estado 3). Las cuentas se revisan cada *mes* y se determina el estado de cada cliente. En general, los créditos no se extienden y se espera que los deudores paguen sus cuentas lo más pronto posible. En ocasiones, los clientes no pagan en la fecha límite. Si esto ocurre cuando el saldo queda dentro de los 30 días de retraso, la tienda considera que este cliente permanece en el estado 1. Si esto ocurre cuando el saldo está entre 31 y 60 días de retraso, la tienda considera que el cliente se mueve al estado 2. Los clientes que tienen más de 60 días de retraso se clasifican en la categoría de una mala deuda (estado 3), en cuyo caso envía las cuentas a una agencia de cobro.

Después de examinar los datos de años anteriores en la progresión mes a mes de los clientes individuales de estado a estado, la tienda ha desarrollado la siguiente matriz de transición:[4]

Estado \ Estado	0: Saldo pagado	1: 1 a 30 días de retraso	2: 31 a 60 días de retraso	3: Mala deuda
0: saldo pagado	1	0	0	0
1: 1 a 30 días de retraso	0.7	0.2	0.1	0
2: 31 a 60 días de retraso	0.5	0.1	0.2	0.2
3: mala deuda	0	0	0	1

Aunque cada cliente acaba por llegar al estado 0 o al estado 3, la tienda se interesa en determinar la probabilidad de que un cliente llegue a ser un mal deudor dado que la cuenta pertenece al estado de 1 a 30 días de retraso, y de igual forma, si se encuentra en 31 a 60 días de retraso.

Para obtener esta información debe resolverse el conjunto de ecuaciones que se presentó al principio de esta sección para obtener f_{13} y f_{23}. Las siguientes dos ecuaciones se obtienen por sustitución:

$$f_{13} = p_{10}f_{03} + p_{11}f_{13} + p_{12}f_{23} + p_{13}f_{33},$$
$$f_{23} = p_{20}f_{03} + p_{21}f_{13} + p_{22}f_{23} + p_{23}f_{33}.$$

Como $f_{03} = 0$ y $f_{33} = 1$, ahora se tienen dos ecuaciones con dos incógnitas, esto es,

$$(1 - p_{11})f_{13} = p_{13} + p_{12}f_{23},$$
$$(1 - p_{22})f_{23} = p_{23} + p_{21}f_{13}.$$

Al sustituir los valores de la matriz de transición se llega a

$$0.8f_{13} = 0.1f_{23},$$
$$0.8f_{23} = 0.2 + 0.1f_{13},$$

y la solución es

$$f_{13} = 0.032,$$
$$f_{23} = 0.254.$$

Entonces, alrededor de 3% de los clientes cuyas cuentas tienen de 1 a 30 días de retraso y alrededor de 25% de los clientes con 31 a 60 días de retraso llegan a ser malos deudores.

16.8 CADENAS DE MARKOV DE TIEMPO CONTINUO

En todas las secciones anteriores se supuso que el parámetro t del tiempo es discreto (es decir, $t = 0, 1, 2, \ldots$). Este supuesto es adecuado para muchos problemas, pero existen ciertos casos

[4] Los clientes que pagaron su saldo (en el estado 0) y después se retrasan en el pago de nuevas compras se consideran clientes “nuevos” que inician en el estado 1.

(como en algunos modelos de colas, que se estudiarán en el capítulo 17) en los que se requiere un parámetro (llamado t') de tiempo continuo, debido a que la evolución del proceso se observa de manera *continua* a través del tiempo. La definición de cadena de Markov que se dio en la sección 16.2 también se extiende a esos procesos continuos. Esta sección está dedicada a la descripción de estas "cadenas de Markov de tiempo continuo" y sus propiedades.

Formulación

Como se hizo con anterioridad, se etiquetan los **estados** posibles del sistema 0, 1, . . ., M. Si se comienza en el tiempo 0 y se deja que el parámetro de tiempo t' corra de manera continua para $t' \geq 0$, sea la variable aleatoria $X(t')$ el estado del sistema en el tiempo t'. Entonces $X(t')$ toma uno de sus $(M + 1)$ valores posibles en un intervalo $0 \leq t' < t_1$, después salta a otro valor en el siguiente intervalo $t_1 \leq t' < t_2$ y así sucesivamente, donde los puntos de tránsito (t_1, t_2, . . .) son puntos aleatorios en el tiempo (*no* necesariamente enteros).

Ahora considere los tres puntos en el tiempo 1) $t' = r$ (donde $r \geq 0$), 2) $t' = s$ (donde $s > r$) y 3) $t' = s + t$ (donde $t > 0$), interpretados como sigue:

$t' = r$ es un tiempo pasado,
$t' = s$ es el tiempo actual,
$t' = s + t$ es t unidades de tiempo hacia el futuro.

Por lo tanto, el estado del sistema se ha observado en los tiempos $t' = s$ y $t' = r$. Estos estados se etiquetan como

$$X(s) = i \qquad \text{y} \qquad X(r) = x(r).$$

Dada esta información, el paso natural es buscar la distribución de probabilidad del estado del sistema en el tiempo $t' = s + t$. En otras palabras, ¿cuál es la probabilidad

$$P\{X(s + t) = j \mid X(s) = i \text{ y } X(r) = x(r)\}, \qquad \text{para } j = 0, 1, \ldots, M?$$

Suele ser muy difícil calcular estas probabilidades condicionales. Sin embargo, esta tarea se simplifica en gran medida si el proceso estocástico involucrado posee la siguiente propiedad clave.

Un proceso estocástico de tiempo continuo $\{X(t'); t' > 0\}$ tiene la **propiedad markoviana** si

$$P\{X(t + s) = j \mid X(s) = i \text{ y } X(r) = x(r)\} = P\{X(t + s) = j \mid X(s) = i\},$$

para toda $i, j = 0, 1, \ldots, M$ y para toda $r \geq 0$, $s > r$, y $t > 0$.

Observe que $P\{X(t + s) = j|X(s) = i\}$ es una **probabilidad de transición**, igual que las probabilidades de transición de las cadenas de Markov de tiempos discretos que se estudió en las secciones anteriores, donde la única diferencia es que ahora no es necesario que t sea entero.

Si las probabilidades de transición son independientes de s, de manera que

$$P\{X(t + s) = j \mid X(s) = i\} = P\{X(t) = j \mid X(0) = i\}$$

para toda $s > 0$, se llaman **probabilidades de transición estacionarias**.

Para simplificar la notación se denotará estas probabilidades estacionarias por

$$p_{ij}(t) = P\{X(t) = j|X(0) = i\},$$

donde se hará referencia a $p_{ij}(t)$ como la **función de probabilidad de transición de tiempo continuo**. Se supone que

$$\lim_{t \to 0} p_{ij}(t) = \begin{cases} 1 & \text{si} \quad i = j \\ 0 & \text{si} \quad i \neq j. \end{cases}$$

Ahora se pueden definir las cadenas de Markov de tiempo continuo que se estudiarán en esta sección.

Un proceso estocástico de tiempo continuo $\{X(t'); t' \geq 0\}$ es una **cadena de Markov de tiempo continuo** si presenta la *propiedad markoviana*.

Recuadro de aplicación

La compañía **PSA Peugeot Citroën**, cuya matriz se encuentra en Francia, es una de las fábricas de autos más grandes del mundo. Cuando se tomó la decisión de introducir 25 nuevos modelos entre 2001 y 2004, la administración de PSA decidió rediseñar sus talleres de carrocerías con el fin de que las de varios modelos pudieran ensamblarse en sólo uno de ellos. Se le asignó a un grupo de IO las tareas de supervisar el proceso de diseño y desarrollar herramientas para evaluar, con antelación, la eficiencia de las líneas de producción de cualquier modelo en los talleres.

El equipo de IO desarrolló métodos aproximados rápidos y métodos detallados más extensos para realizar esta tarea. Sin embargo, un factor clave que debía ser incorporado en todos los métodos era determinar la frecuencia con que cada una de las máquinas del taller se descomponía y, por lo tanto, debía ser reparada, lo cual detenía el flujo de trabajo en ese taller. El grupo de IO utilizó una *cadena de Markov de tiempo continuo* para representar la evolución de cada tipo de máquina en el cambio que representa que la máquina esté en estado operativo (que funcione correctamente) y sin funcionar (descompuesta). Por lo tanto, la cadena de Markov sólo tiene dos estados, *en funcionamiento* y *descompuesta*, con un periodo (pequeño) de transición entre los estados en funcionamiento y descompuesta y un periodo de transición (más largo) entre los estados *descompuesta* y *en funcionamiento*. El equipo llegó a la conclusión de que la velocidad de transición entre ambos estados es esencialmente el mismo sin tomar en cuenta si la máquina en realidad trabaja o está ociosa, por lo que no es necesario dividir el primero de ellos en los estados en *operación* y *ocioso*.

Esta aplicación de la IO, que incluyó esta sencilla cadena de Markov de tiempo continuo, tuvo un efecto dramático en la compañía. Mediante una mejora sustancial de la eficiencia de las líneas de producción en los talleres PSA con una mínima inversión de capital y sin comprometer la calidad, dicha aplicación *contribuyó* con **130 millones de dólares** *a las ganancias de PSA* (alrededor de 6.5% de la ganancia total) sólo en el primer año.

Fuente: A. Patchong, T. Lemoine y G. Kern: "Improving Car Body Production at PSA Peugeot Citroën", en *Interfaces*, **33**(1): 36-49, enero-febrero de 2003. (En el sitio en internet de este libro —www.mhhe.com/hillier—) se proporciona una liga hacia este artículo.)

Aquí se restringirá el estudio a las cadenas de Markov de tiempo continuo con las siguientes propiedades.

1. Un número finito de estados.
2. Probabilidades de transición estacionarias.

Algunas variables aleatorias importantes

En el análisis de las cadenas de Markov de tiempo continuo, un conjunto importante de variables aleatorias es el siguiente.

> Cada vez que el proceso entra en el estado i, la cantidad de tiempo que pasa en ese estado antes de moverse a uno diferente es una variable aleatoria T_i, donde $i = 0, 1, \ldots, M$.

Suponga que el proceso entra en el estado i en el tiempo $t' = s$. Entonces, para cualquier cantidad de tiempo fijo $t > 0$, observe que $T_i > t$ si y sólo si $X(t') = i$ para toda t' en el intervalo $s \le t' \le s + t$. Por lo tanto, la propiedad markoviana (con probabilidades de transición estacionarias) implica que

$$P\{T_i > t + s \mid T_i > s\} = P\{T_i > t\}.$$

Ésta es una condición bastante rara para una distribución de probabilidad. Dice que la distribución de probabilidad del tiempo *que falta* para que el proceso haga una transición fuera de un estado dado siempre es la misma, independientemente de cuánto tiempo haya pasado el proceso en ese estado. En efecto, la variable aleatoria no tiene memoria; el proceso olvida su historia. Existe sólo una distribución de probabilidad (continua) que posee esta propiedad, la *distribución exponencial*. Esta distribución tiene un solo parámetro, llámese q, donde la media es $1/q$ y la función de distribución acumulada es

$$P\{T_i \le t\} = 1 - e^{-qt}, \qquad \text{para } t \ge 0.$$

(En la sección 17.4 se describirá con detalle las propiedades de la distribución exponencial.)

Este resultado conduce a una forma equivalente para describir una cadena de Markov de tiempo continuo:

1. La variable aleatoria T_i tiene una distribución exponencial con media $1/q_i$.
2. Cuando sale de un estado i, el proceso se mueve a otro estado j, con probabilidad p_{ij}, donde p_{ij} satisface las condiciones

$$p_{ii} = 0 \qquad \text{para toda } i,$$

y

$$\sum_{j=0}^{M} p_{ij} = 1 \qquad \text{para toda } i.$$

3. El siguiente estado que se visita después del estado i es independiente del tiempo que pasó en el estado i.

Igual que las probabilidades de transición de un paso tuvieron un papel primordial para describir una cadena de Markov de tiempos discretos, el papel análogo en el caso de la cadena de Markov de tiempo continuo lo tienen las intensidades de transición.

Las **intensidades de transición** son

$$q_i = -\frac{d}{dt}p_{ii}(0) = \lim_{t\to 0} \frac{1 - p_{ii}(t)}{t}, \qquad \text{para } i = 0, 1, 2, \ldots, M,$$

y

$$q_{ij} = \frac{d}{dt}p_{ij}(0) = \lim_{t\to 0} \frac{p_{ij}(t)}{t} = q_i p_{ij}, \qquad \text{para todo } j \neq i,$$

donde $p_{ij}(t)$ es la *función de probabilidad de transición de tiempo continuo* que se presentó al principio de la sección y p_{ij} es la probabilidad descrita en la propiedad 2 del párrafo anterior. Más aún, q_i, según se definió aquí, también resulta ser el parámetro de la distribución exponencial de T_i (vea la propiedad 1 del párrafo anterior).

La interpretación intuitiva de q_i y q_{ij} es que son *tasas de transición*. En particular, q_i es la *tasa de transición hacia fuera del estado i* en el sentido de que q_i es el número esperado de veces que el proceso deja el estado i por unidad de tiempo que pasa en el estado i. (De esta forma, q_i es el recíproco del tiempo esperado que el proceso pasa en el estado i por cada visita al estado i; es decir, $q_i = 1/E[T_i]$.) De manera similar, q_{ij} es la *tasa de transición del estado i al estado j* en el sentido de que q_{ij} es el número esperado de veces que el proceso transita del estado i al estado j por unidad de tiempo que pasa en el estado i. Así,

$$q_i = \sum_{j\neq i} q_{ij}.$$

Igual que q_i es el parámetro de la distribución exponencial de T_i, cada q_{ij} es el parámetro de una distribución exponencial de una variable aleatoria relacionada que se describe en seguida.

Cada vez que el proceso entra al estado i, la cantidad de tiempo que pasará en el estado i antes de que ocurra una transición al estado j (si no ocurre antes una transición a algún otro estado) es una variable aleatoria T_{ij}, donde $i, j = 0, 1, \ldots, M$ y $j \neq i$. Las T_{ij} son variables aleatorias independientes, donde cada T_{ij} tiene una *distribución exponencial* con parámetro q_{ij}, de manera que $E[T_{ij}] = 1/q_{ij}$. El tiempo que pasa en el estado i hasta que ocurre una transición (T_i) es el *mínimo* (sobre $j \neq i$) de las T_{ij}. Cuando ocurre la transición, la probabilidad de que sea al estado j es $p_{ij} = q_{ij}/q_i$.

Probabilidades de estado estable

Igual que las probabilidades de transición de una cadena de Markov de tiempos discretos satisfacen las ecuaciones de Chapman-Kolmogorov, la función de probabilidad de transición de tiempo con-

tinuo satisface estas ecuaciones. Entonces, para cualesquiera estados i y j, y números no negativos t y s $(0 \leq s \leq t)$,

$$p_{ij}(t) = \sum_{k=0}^{M} p_{ik}(s)p_{kj}(t - s).$$

Se dice que un par de estados i y j *se comunican* si existen tiempos t_1 y t_2 tales que $p_{ij}(t_1) > 0$ y $p_{ji}(t_2) > 0$. Se dice que todos los estados que se comunican forman una *clase*. Si todos los estados de una cadena forman una sola clase, es decir, si la cadena de Markov es *irreducible* (lo que se supondrá de aquí en adelante), entonces,

$$p_{ij}(t) > 0, \qquad \text{para toda } t > 0 \text{ y todos los estados } i \text{ y } j.$$

Aún más,

$$\lim_{t \to \infty} p_{ij}(t) = \pi_j$$

siempre existe y es independiente del estado inicial de la cadena de Markov, para $j = 0, 1, \ldots, M$. Estas probabilidades limitantes se conocen como las **probabilidades de estado estable** (o *probabilidades estacionarias*) de la cadena de Markov.

Las π_j satisfacen las ecuaciones

$$\pi_j = \sum_{i=0}^{M} \pi_i p_{ij}(t), \qquad \text{para } j = 0, 1, \ldots, M \text{ y para toda } t \geq 0.$$

Sin embargo, las siguientes **ecuaciones de estado estable** proporcionan un sistema de ecuaciones más útil para obtener las probabilidades de estado estable:

$$\pi_j q_j = \sum_{i \neq j} \pi_i q_{ij}, \qquad \text{para } j = 0, 1, \ldots, M.$$

y

$$\sum_{j=0}^{M} \pi_j = 1.$$

La ecuación de estado estable del estado j tiene una interpretación intuitiva. El lado izquierdo $(\pi_j q_j)$ es la *tasa* a la que el proceso *deja* el estado j, puesto que π_j es la probabilidad (estable) de que el proceso esté en el estado j y q_j es la tasa de transición hacia fuera de j dado que el proceso se encuentra en el estado j. De manera similar, cada término del lado derecho $(\pi_j q_{ij})$ es la *tasa* a la que el proceso *entra* al estado j desde el estado i, puesto que q_{ij} es la tasa de transición del estado i al j dado que el proceso se encuentra en el estado i. Si se suma sobre toda $i \neq j$, todo el lado derecho proporciona la tasa a la que el proceso entra al estado j desde cualquier otro estado. Por eso la ecuación global establece que la tasa a la cual el proceso deja el estado j debe ser igual a la tasa a la que entra en el estado j. Así, esta ecuación es análoga a las ecuaciones de conservación del flujo de los cursos de ingeniería y ciencias.

Como cada una de las primeras $M + 1$ *ecuaciones de estado estable* requiere que las dos tasas *estén balanceadas* (sean iguales), a veces estas ecuaciones se llaman **ecuaciones de balance**.

Ejemplo. Un taller tiene dos máquinas idénticas en operación continua excepto cuando se descomponen. Como lo hacen con bastante frecuencia, la tarea con más alta prioridad para la persona de mantenimiento que trabaja tiempo completo es repararlas cuando sea necesario.

El tiempo que se requiere para reparar una máquina tiene distribución exponencial con media de $\frac{1}{2}$ día. Una vez que se termina la reparación, el tiempo que transcurre hasta la siguiente descompostura tiene distribución exponencial con media de un día. Estas distribuciones son independientes.

Defina la variable aleatoria $X(t')$ como

$$X(t') = \text{número de máquinas descompuestas en el tiempo } t',$$

de forma que los valores posibles de $X(t')$ son 0, 1, 2. Por lo tanto, si se corre el parámetro t' de manera continua desde el tiempo 0, el proceso estocástico de tiempo continuo $\{X(t'); t' \geq 0\}$ proporciona la evolución del número de máquinas descompuestas.

Como tanto el tiempo de reparación como el tiempo hasta la siguiente descompostura tienen distribuciones exponenciales, $\{X(t'); t' \geq 0\}$ es una *cadena de Markov de tiempo continuo*[5] con estados 0, 1, 2. En consecuencia, se pueden usar las probabilidades de estado estable dadas en la subsección anterior para encontrar la distribución de probabilidad de estado estable del número de máquinas descompuestas. Para hacer esta tarea se debe determinar todas las *tasas de transición*, esto es, las q_i y q_{ij} para $i, j = 0, 1, 2$.

El estado (número de máquinas descompuestas) aumenta en 1 cuando ocurre una descompostura y disminuye en 1 cuando se termina una reparación. Como tanto las descomposturas como las reparaciones ocurren una a la vez, $q_{02} = 0$ y $q_{20} = 0$. El tiempo esperado de reparación es de $\frac{1}{2}$ día, de manera que la tasa a la que se terminan las reparaciones (cuando hay máquinas descompuestas) es 2 por día, lo que implica que $q_{21} = 2$ y $q_{10} = 2$. De manera similar, el tiempo esperado hasta que se descompone una máquina en operación es de un día, de manera que la tasa a la que se descompone (cuando está en operación) es de un por día; esto implica que $q_{12} = 1$. Durante los tiempos en los que las dos máquinas operan, las descomposturas ocurren a una tasa de $1 + 1 = 2$ por día, por lo que $q_{01} = 2$.

Estas tasas de transición se resumen en el diagrama de tasas que se muestra en la figura 16.5. Se pueden usar para calcular la *tasa de transición total* hacia fuera de cada estado.

$$q_0 = q_{01} = 2.$$
$$q_1 = q_{10} + q_{12} = 3.$$
$$q_2 = q_{21} = 2.$$

Si se sustituyen todas las tasas en las ecuaciones de estado estable de la subsección anterior, se obtiene:

Ecuación de balance para el estado 0: $2\pi_0 = 2\pi_1$
Ecuación de balance para el estado 1: $3\pi_1 = 2\pi_0 + 2\pi_2$
Ecuación de balance para el estado 2: $2\pi_2 = \pi_1$
Las probabilidades suman 1: $\pi_0 + \pi_1 + \pi_2 = 1$

Cualquiera de las ecuaciones de balance (por ejemplo, la segunda) se puede eliminar como redundante, y la solución simultánea de las ecuaciones restantes proporciona la distribución de estado estable como

$$(\pi_0, \pi_1, \pi_2) = \left(\frac{2}{5}, \frac{2}{5}, \frac{1}{5}\right).$$

Entonces, en el largo plazo, ambas máquinas estarán descompuestas en forma simultánea 20% del tiempo y una máquina estará descompuesta otro 40% del tiempo.

En el siguiente capítulo (sobre teoría de colas) se incluyen muchos ejemplos de cadenas de Markov de tiempo continuo. En realidad, la mayor parte de los modelos básicos de la teoría

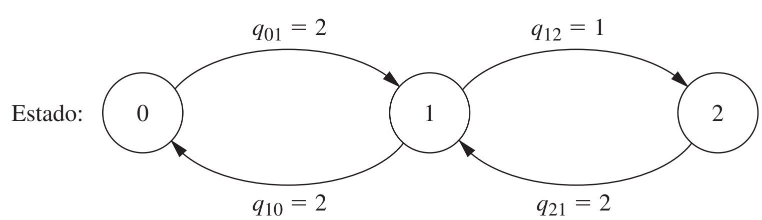


FIGURA 16.5
Diagrama de tasas del ejemplo de una cadena de Markov de tiempo continuo.

[5] La prueba de este hecho necesita usar dos propiedades de la distribución exponencial que se presentan en la sección 17.4 (la *falta de memoria* y que el *mínimo de exponenciales es exponencial*), puesto que estas propiedades implican que las variables aleatorias T_{ij} que se introdujeron antes sin duda tienen distribución exponencial.

de colas cae dentro de esta categoría. El ejemplo que se acaba de dar en realidad se ajusta a uno de estos modelos (la variación de fuente de entrada finita al modelo *M/M/s* que se incluye en la sección 17.6).

REFERENCIAS SELECCIONADAS

1. Bhat, U. N. y G. K. Miller: *Elements of Applied Stochastic Processes*, 3a. ed., Wiley, Nueva York, 2002.
2. Bini, D., G. Latouche y B. Meini: *Numerical Methods for Structured Markov Chains*, Oxford University Press, Nueva York, 2005.
3. Bukiet, B., E. R. Harold y J. L. Palacios: "A Markov Chain Approach to Baseball", en *Operations Research*, **45:** 14-23, 1997.
4. Ching, W. K. y M. K. Ng: *Markov Chains: Models, Algorithms and Applications*, Springer, Nueva York, 2006.
5. Grassmann, W. K. (ed.) *Computational Probability*; Kluwer Academic Publishers (actualmente Springer), Boston, MA, 2000.
6. Mamon, R. S. y R. J. Elliot (eds.): *Hidden Markov Models in Finance*, Springer, Nueva York, 2007.
7. Resnick, S. I.: *Adventures in Stochastic Processes*, Birkhäuser, Boston, 1992.
8. Tijms, H. C.: *A First Course in Stochastic Models*, Wiley, Nueva York, 2003.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos para el capítulo 16

Rutinas automáticas en IOR Tutorial:

Introducción de una matriz de transición (Enter Transition Matrix)
Ecuaciones Chapman-Kolmogorov (Chapman-Kolmogorov Equations)
Probabilidades de estado estable (Steady-State Probabilities)

Archivo de LINGO para ejemplos seleccionados "Ch. 16—Markov Chains"

Glosario para el capítulo 16

Vea el apéndice 1 para la documentación del software.

PROBLEMAS

Los símbolos a la izquierda de algunos problemas (o de sus incisos) significan lo siguiente:

C: Use la computadora con las rutinas automáticas correspondientes mencionadas con anterioridad (u otras equivalentes) para resolver el problema.

Un asterisco en el número del problema indica que al final del libro se da al menos una respuesta parcial.

16.2-1. Lea el artículo de referencia que describe el estudio de IO que se resume en el recuadro de aplicación que se presentó en la sección 16.2. Describa de manera breve la forma en que se aplicó una cadena de Markov a este estudio. Después, elabore una lista de los beneficios financieros y de otro tipo que arrojó dicho estudio.

16.2-2. Suponga que la probabilidad de lluvia mañana es de 0.5 si hoy llueve y que la probabilidad de un día claro (sin lluvia) mañana es de 0.9 si hoy está despejado. Suponga además que estas probabilidades no cambian si también se proporciona información sobre el clima de días anteriores a hoy.

a) Explique por qué los supuestos establecidos implican que la *propiedad markoviana* se cumple en el caso de la evolución del clima.

b) Formule la evolución del clima como una cadena de Markov mediante la definición de sus estados y la construcción de su matriz de transición (de un paso).

16.2-3. Considere la segunda versión del modelo de mercado de acciones que se presentó en la sección 16.2. Si mañana la acción

sube o no depende de si subió o no hoy *y* ayer. Si la acción subió hoy y ayer, mañana subirá con probabilidad α_1. Si la acción subió hoy y ayer bajó, mañana subirá con probabilidad α_2. Si la acción bajó hoy y ayer subió, la probabilidad de que suba mañana es α_3. Por último, si la acción bajó hoy y ayer, la probabilidad de que suba mañana es α_4.

a) Construya la matriz de transición de un paso para la cadena de Markov.

b) Explique por qué los estados que se utilizaron para esta cadena de Markov hacen que la definición matemática de la propiedad markoviana se cumpla aunque lo que pase en el futuro (mañana) depende de lo que ocurrió en el pasado (ayer) y en el presente (hoy).

16.2-4. Reconsidere el problema 16.2-3. Suponga ahora que el hecho de que la acción suba mañana depende de si subió o no hoy, ayer *y* anteayer. ¿Puede este problema formularse como una cadena de Markov? Si se puede, ¿cuáles son los estados posibles? Explique por qué estos estados dan al proceso la *propiedad markoviana*, mientras que los estados del problema 16.2-3 no lo hacen.

16.3-1. Reconsidere el problema 16.2-2.

C ***a***) Utilice la rutina de las *ecuaciones de Chapman-Kolmogorov* del IOR Tutorial para encontrar la matriz de transición de n pasos $\mathbf{P}^{(n)}$ para $n = 2, 5, 10, 20$.

b) La probabilidad de que llueva hoy es de 0.5. Use los resultados del inciso *a*) para determinar la probabilidad de que llueva dentro de n días, para $n = 2, 5, 10, 20$.

C ***c***) Utilice la rutina de *probabilidades de estado estable* del IOR Tutorial para determinar las probabilidades de estado estable del clima. Compare las probabilidades de las matrices de transición de n pasos del inciso *a*) compare estas probabilidades de estado estable conforme n crece.

16.3-2. Suponga que una red de comunicaciones transmite dígitos binarios, 0 o 1, y que cada dígito se transmite 10 veces sucesivas. Durante cada transmisión, la probabilidad de que ese dígito se transmita correctamente es de 0.995. En otras palabras, existe una probabilidad de 0.005 de que el dígito transmitido se registre con el valor opuesto al final de la transmisión. Para cada transmisión después de la primera, el dígito transmitido es el que se registra al final de la transmisión anterior. Si X_0 denota el dígito binario que entra al sistema, X_1 el dígito binario que se apunta después de la primera transmisión, X_2 el dígito binario que se anota después de la segunda transmisión, . . ., entonces $\{X_n\}$ es una cadena de Markov.

a) Determine la matriz de transición (de un paso).

C ***b***) Utilice el IOR Tutorial para encontrar la matriz de transición de 10 pasos $\mathbf{P}^{(10)}$. Utilice este resultado para identificar la probabilidad de que un dígito que entra a la red se registre correctamente después de la última transmisión.

C ***c***) Suponga que la red se rediseña para mejorar la probabilidad de la exactitud de una sola transmisión de 0.995 a 0.998. Repita el inciso *b*) para encontrar la nueva probabilidad de que un dígito que entra a la red se registre correctamente después de la última transmisión.

16.3-3.* Una partícula se mueve sobre un círculo por puntos marcados 0, 1, 2, 3, 4 (en el sentido de las manecillas del reloj). La partícula comienza en el punto 0. En cada paso tiene una probabilidad de 0.5 de moverse un punto en el sentido de las manecillas del reloj (0 sigue al 4) y una probabilidad de 0.5 de moverse un punto en el sentido opuesto. Sea X_n ($n \geq 0$) la localización en el círculo después del paso n. $\{X_n\}$ es entonces una cadena de Markov.

a) Encuentre la matriz de transición (de un paso).

C ***b***) Utilice el IOR Tutorial para determinar la matriz de transición de n pasos $\mathbf{P}^{(n)}$ para $n = 5, 10, 20, 40, 80$.

C ***c***) Utilice el IOR Tutorial para determinar las probabilidades de estado estable de los estados de la cadena de Markov. Compare las probabilidades de la matriz de transición de n pasos que se obtuvo en el inciso *b*) con estas probabilidades de estado estable conforme n crece.

16.4-1.* Dadas las siguientes matrices de transición (de un paso) de una cadena de Markov, determine las clases de las cadenas de Markov y si son recurrentes o no.

$$a)\quad \mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

$$b)\quad \mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix} \end{array}$$

16.4-2. Dadas las siguientes matrices de transición (de un paso) de una cadena de Markov, determine las clases de las cadenas de Markov y si son recurrentes o no.

$$a)\quad \mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix} \end{array}$$

$$b)\quad \mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \end{bmatrix} \end{array}$$

16.4-3. Dada la siguiente matriz de transición (de un paso), determine las clases de la cadena de Markov y si son recurrentes o no.

$$\mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 \\ \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{bmatrix} \end{array}$$

16.4-4. Determine el periodo de los estados de la cadena de Markov que tiene la siguiente matriz de transición (de un paso).

$$\mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & \frac{3}{4} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \end{bmatrix} \end{array}$$

16.4-5. Considere la cadena de Markov que tiene la siguiente matriz de transición (de un paso).

$$\mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & \frac{4}{5} & 0 & \frac{1}{5} & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{10} & \frac{2}{5} \\ 0 & 0 & 0 & 1 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix} \end{array}$$

a) Determine las clases de esta cadena de Markov y, para cada clase, determine si es recurrente o transitoria.

b) Para cada una de las clases identificadas en el inciso *a*), determine el periodo de los estados de esa clase.

16.5-1. Reconsidere el problema 16.2-2. Suponga ahora que las probabilidades dadas, 0.5 y 0.9, se sustituyen por valores arbitrarios α y β, respectivamente. Obtenga las *probabilidades de estado estable* del estado del clima en términos de α y β.

16.5-2. Se dice que una matriz de transición **P** es doblemente estocástica si la suma de los elementos de cada columna es igual a 1; esto es,

$$\sum_{i=0}^{M} p_{ij} = 1, \quad \text{para toda } j.$$

Si esta cadena es irreducible y aperiódica y consiste en $M + 1$ estados, demuestre que

$$\pi_j = \frac{1}{M+1}, \quad \text{para } j = 0, 1, \ldots, M.$$

16.5-3. Reconsidere el problema 16.3-3. Use los resultados del problema 16.5-2 para encontrar las probabilidades de estado estable de esta cadena de Markov. Luego encuentre qué pasa con estas probabilidades si, en cada paso, la probabilidad de moverse a un punto en el sentido de las manecillas del reloj cambia a 0.9 y la probabilidad de moverse a un punto en sentido opuesto cambia a 0.1.

C **16.5-4.** La cervecería más importante de la costa oeste (denotada con la letra *A*) ha contratado a un experto en IO para que analice su posición en el mercado. En especial, la empresa está preocupada por las actividades de su mayor competidor (denotada con la letra *B*). El analista piensa que el cambio de marca se puede modelar como una cadena de Markov que incluya tres estados: los estados *A* y *B* representan a los clientes que beben cerveza que producen las mencionadas cervecerías y el estado *C* representa todas las demás marcas. Los datos se toman cada mes y el analista construye la siguiente matriz de transición (de un paso) con datos históricos.

	A	*B*	*C*
A	0.7	0.2	0.1
B	0.2	0.75	0.05
C	0.1	0.1	0.8

¿Cuáles son los porcentajes de mercado en el estado estable de las dos cervecerías grandes?

16.5-5. Considere el siguiente problema de inventario de sangre al que se enfrenta un hospital. Se tiene necesidad de un tipo raro de sangre, como AB, Rh negativo. La demanda *D* (en pintas) durante un periodo de tres días está dada por

$$P\{D = 0\} = 0.4, \qquad P\{D = 1\} = 0.3,$$
$$P\{D = 2\} = 0.2, \qquad P\{D = 3\} = 0.1.$$

Observe que la demanda esperada es de una pinta, puesto que $E(D) = 0.3(1) + 0.2(2) + 0.1(3) = 1$. Suponga que se surte sangre cada tres días. El hospital propone una política para recibir una pinta en cada entrega y usar primero la más antigua. Si se requiere más sangre de la que hay en el banco se hace un pedido de emergencia a un alto costo. La sangre se descarta si en 21 días no se ha usado. Denote el estado del sistema como el número de pintas en inventario exactamente después de una entrega. Observe que debido a la política de descartar la sangre, el estado más grande posible es 7.

a) Construya la matriz de transición (de un paso) para esta cadena de Markov.

C ***b***) Encuentre las probabilidades de estado estable para los estados de esta cadena de Markov.

c) Use los resultados de *b*) para encontrar la probabilidad de estado estable de que sea necesario descartar una pinta durante un periodo de tres días. (*Sugerencia:* Si se usa primero la sangre más vieja, una pinta tiene 21 días sólo si el estado es 7 y entonces $D = 0$.)

d) Utilice los resultados de *b*) para encontrar la probabilidad de estado de que se necesite una entrega de emergencia durante los tres días entre entregas normales.

C **16.5-6.** En la sección 16.5 se calculó el costo promedio esperado (a largo plazo) por semana (basado sólo en costos de ordenar y costos de la demanda insatisfecha) del ejemplo del inventario de cámaras de la sección 16.1. Suponga que se cambia la política de inventarios. Siempre que el número de cámaras al final de la semana sea 0 o 1, se coloca una orden que aumente este número hasta 3. De otra manera, no se coloca una orden.

Calcule de nuevo el costo promedio esperado (a largo plazo) por semana con esta nueva política de inventarios.

16.5-7.* Considere el ejemplo de inventarios de la sección 16.1 con el siguiente cambio en la política. Si el nivel del inventario al final de la semana es 0 o 1, se ordenan dos cámaras adicionales. De otra manera, no se ordena ninguna. Suponga que los costos de almacenamiento son los mismos que en la segunda subsección de la sección 16.5.

C *a*) Encuentre las probabilidades de estado estable de los estados de esta cadena de Markov.

b) Encuentre el costo de almacenamiento esperado a largo plazo por semana.

16.5-8. Considere la siguiente política de inventarios para cierto producto. Si la demanda durante un periodo excede el número de unidades disponibles, esta demanda insatisfecha se considera como un faltante y se surte cuando se recibe la siguiente orden. Sea Z_n ($n = 0, 1, \ldots$) la cantidad de inventario menos el número de unidades faltantes antes de ordenar al final del siguiente periodo n ($Z_0 = 0$). Si Z_n es cero o positivo, no hay pedidos atrasados. Si Z_n es negativo, entonces $-Z_n$ representa el número de unidades faltantes y el inventario es 0. Si al final del periodo n, $Z_n < 1$, se hace un pedido de $2m$ unidades, donde m es el entero más pequeño tal que $Z_n + 2m \geq 1$. Las órdenes se surten de inmediato.

Sean $D_1, D_2, \ldots$, las demandas respectivas de un producto en los periodos 1, 2, . . . Suponga que las D_n son variables aleatorias, independientes e idénticamente distribuidas que toman los valores 0, 1, 2, 3, 4, cada uno con probabilidad $\frac{1}{5}$. Sea X_n la cantidad en inventario *después* de ordenar al final del periodo n (con $X_0 = 2$), entonces

$$X_n = \begin{cases} X_{n-1} - D_n + 2m & \text{si } X_{n-1} - D_n < 1 \\ X_{n-1} - D_n & \text{si } X_{n-1} - D_n \geq 1 \end{cases} \qquad (n = 1, 2, \ldots),$$

donde $\{X_n\}$ ($n = 0, 1, \ldots$) es una cadena de Markov. Tiene sólo dos estados, 1 y 2, ya que sólo se ordena cuando $Z_n = 0, -1, -2$ o -3, en cuyo caso se pide 2, 2, 4 y 4 unidades, con lo que queda $X_n = 2, 1, 2, 1$, respectivamente.

a) Encuentre la matriz de transición (de un paso).

b) Use las ecuaciones de estado estable para determinar a mano las probabilidades de estado estable.

c) Ahora use el resultado del problema 16.5-2 para encontrar las probabilidades de estado estable.

d) Suponga que el costo de ordenar es $(2 + 2m)$ si se coloca una orden, y 0 en el caso contrario. El costo del inventario por periodo es Z_n si $Z_n \geq 0$ y cero en otro caso. El costo por incumplimiento en cada periodo es $-4Z_n$ si $Z_n < 0$ y cero en el otro caso. Encuentre el costo promedio esperado (a largo plazo) por unidad de tiempo.

16.5-9 Una unidad importante consta de dos componentes colocadas en paralelo. La unidad tiene un desempeño satisfactorio si una de las dos componentes está en operación. Por lo tanto, sólo se opera una de ellas a la vez, pero ambas se mantienen operativas (capaces de operar) tanto como sea posible, reparándolas cuando se necesite. Un componente operativo tiene una probabilidad de 0.2 de descomponerse en un periodo dado. Cuando ocurre, el componente en paralelo opera, si está operativo, al comenzar el siguiente periodo. Sólo se puede reparar un componente a la vez. Una reparación se inicia al principio del primer periodo disponible y termina al final del siguiente. Sea X_t un vector con dos elementos U y V, donde U es el número de componentes operativos al final del periodo t y V el número de periodos de reparación que transcurren para componentes que todavía no son operativos. Entonces, $V = 0$ si $U = 2$ o si $U = 1$ y la reparación del componente no operativo se está realizando. Como la reparación toma dos periodos, $V = 1$ si $U = 0$ (pues el componente no operativo espera iniciar su reparación mientras la otra entra al segundo periodo) o si $U = 1$ y el componente no operativo está en su segundo periodo. Así, el espacio de estados contiene cuatro estados (2, 0), (1, 0), (0, 1) y (l, 1). Denote estos estados por 0, 1, 2, 3, respectivamente. $\{X_t\}$ ($t = 0, 1, \ldots$) es una cadena de Markov (suponga que $X_0 = 0$) con matriz de transición (de un paso)

$$\mathbf{P} = \begin{array}{c} \text{Estado} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \\ 0 & 1 & 0 & 0 \\ 0.8 & 0.2 & 0 & 0 \end{bmatrix} \end{array}.$$

C *a*) ¿Cuál es la probabilidad de que la unidad no esté operable después de n periodos (porque ambas componentes estén descompuestas), para $n = 2, 5, 10, 20$?

C *b*) ¿Cuáles son las probabilidades de estado estable del estado de esta cadena de Markov?

c) Si cuesta 30 000 dólares por periodo que la unidad no opere (ambas componentes descompuestas) y cero en otro caso, ¿cuál es el costo promedio esperado (a la larga) por periodo?

16.6-1. Una computadora se inspecciona cada hora. Se encuentra que está trabajando o que está descompuesta. En el primer caso, la probabilidad de que siga así la siguiente hora es de 0.95. Si está descompuesta, se repara, lo que puede llevar más de una hora. Siempre que la computadora esté descompuesta (sin importar cuánto tiempo pase), la probabilidad de que siga descompuesta una hora más es de 0. 5.

a) Construya la matriz de transición de un paso de esta cadena de Markov.

b) Utilice el enfoque descrito en la sección 16.6 para encontrar las μ_{ij} (el tiempo esperado de primera pasada del estado i al estado j) para toda i y j.

16.6-2. Un fabricante tiene una máquina que cuando empieza a operar al inicio del día tiene una probabilidad de 0.1 de descomponerse en algún momento de ese día. Cuando esto ocurre, la reparación se hace al siguiente día y se termina al finalizar ese día.

a) Formule la evolución del estado de la máquina como una cadena de Markov; identifique los tres estados posibles al final del día y después construya la matriz de transición (de un paso).

b) Utilice el enfoque descrito en la sección 16.6 para encontrar las μ_{ij} (tiempo esperado de primera pasada del estado i al estado j) para toda i y j. Use estos resultados para identificar la siguiente descompostura después de que se ha terminado una reparación.

c) Ahora suponga que la máquina tiene ya 20 días sin descomponerse desde la última reparación. Compare el número esperado de días completos que, *en adelante*, la máquina permanecerá en operación antes de la siguiente descompostura con el resultado correspondiente del inciso *b*) cuando se acaba de completar una reparación. Proporcione una explicación.

16.6-3. Reconsidere el problema 16.6-2. Suponga que el fabricante tiene una máquina de repuesto que usa sólo cuando la máquina principal está en reparación. Durante el día de reparación, la máquina de repuesto tiene una probabilidad de 0.1 de descomponerse, en cuyo caso se repara el siguiente día. Denote los estados del sistema por (x, y), donde x y y toman valores 1 o 0 según si la primera máquina (x) y la máquina de repuesto (y) están en operación (valor 1) o no (valor 0) al final del día. [*Sugerencia*: observe que (0, 0) no es un estado posible.]

a) Construya la matriz de transición (de un paso) de esta cadena de Markov.
b) Encuentre el *tiempo esperado de recurrencia* del estado (1, 0).

16.6-4. Considere el ejemplo de inventarios de la sección 16.1 pero ahora la demanda tiene la siguiente distribución de probabilidad:

$$P\{D = 0\} = \frac{1}{4}, \qquad P\{D = 2\} = \frac{1}{4},$$

$$P\{D = 1\} = \frac{1}{2}, \qquad P\{D \geq 3) = 0.$$

La política de órdenes cambia al ordenar exactamente dos cámaras al final de la semana si el inventario es cero. Como antes, no se ordena si hay cámaras en inventario. Suponga que se tiene una cámara (al final de la semana) cuando se instituye la política.

a) Construya la matriz de transición (de un paso).
C ***b***) Encuentre la distribución de probabilidad del estado de esta cadena de Markov n semanas después de instituir la nueva política de inventarios, para $n = 2, 5, 10$.
c) Encuentre las μ_{ij} (el tiempo esperado de primera pasada del estado i al j) para toda i y j.
C ***d***) Encuentre las probabilidades de estado estable del estado de esta cadena de Markov.
e) Bajo el supuesto de que la tienda paga un costo de almacenamiento por cada cámara que queda en la repisa al final de la semana, de acuerdo con la función $C(0) = 0$, $C(1) = \$2$ y $C(2) = \$8$, encuentre el costo promedio de almacenamiento a largo plazo por semana.

16.6-5. Un proceso de producción incluye una máquina que se deteriora con rapidez tanto en la calidad como en la cantidad de producción con el trabajo pesado, por lo que se inspecciona al final de cada día. Después de la inspección se clasifica la condición de la máquina en uno de cuatro estados posibles:

Estado	Condición
0	Tan buena como nueva
1	Operable: deterioro mínimo
2	Operable: deterioro mayor
3	Inoperable y reemplazada por una tan buena como nueva

El proceso se puede modelar como una cadena de Markov con matriz de transición (de un paso) **P** dada por

Estado	0	1	2	3
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	0	$\frac{1}{2}$	$\frac{1}{2}$
3	1	0	0	0

C ***a***) Encuentre las probabilidades de estado estable.
b) Si los costos respectivos por estar en los estados 0, 1, 2, 3 son 0, 1 000, 3 000 y 6 000 dólares, ¿cuál es el costo diario esperado a largo plazo?
c) Encuentre el *tiempo de recurrencia esperado* del estado 0 (esto es, el tiempo esperado que una máquina se puede usar antes de tener que reemplazarla).

16.7-1. Considere el siguiente problema de la ruina de un jugador. Un jugador apuesta 1 dólar en cada jugada. Tiene una probabilidad p de ganar y $q = 1 - p$ de perder. Continuará jugando hasta que quiebre o reúna una fortuna de T dólares. Sea X_n la fortuna del jugador en la n-ésima jugada. Entonces,

$$X_{n+1} = \begin{cases} X_n + 1 & \text{con probabilidad } p \\ X_n - 1 & \text{con probabilidad } q = 1 - p \end{cases} \qquad \text{para } 0 < X_n < T,$$

$$X_{n+1} = X_n, \qquad \text{para } X_n = 0, \text{ o } T.$$

$\{X_n\}$ es una cadena de Markov. El jugador comienza con X_0 dólares, donde X_0 es un entero positivo menor que T.

a) Determine la matriz de transición (de un paso) de la cadena de Markov.
b) Encuentre las clases de la cadena de Markov.
c) Sea $T = 3$ y $p = 0.3$. Use la notación de la sección 16.7 para encontrar $f_{10}, f_{1T}, f_{20}, f_{2T}$.
d) Sea $T = 3$ y $p = 0.7$. Encuentre $f_{10}, f_{1T}, f_{20}, f_{2T}$.

16.7-2. Un fabricante de videograbadoras está tan seguro de su calidad que ofrece garantía de reposición total si un aparato falla dentro de los dos primeros años. Con base en datos compilados, la compañía ha notado que sólo 1% de sus grabadoras fallan durante el primer año, mientras que 5% de ellas sobreviven el primer año pero fallan durante el segundo. La garantía no cubre grabadoras ya reemplazadas.

a) Formule la evolución del estado de una grabadora como una cadena de Markov cuyos estados incluyen dos estados absorbentes que representan la necesidad de cubrir la garantía o el hecho de que una grabadora sobreviva el periodo de garantía. Después construya la matriz de transición (de un paso).
b) Utilice el enfoque descrito en la sección 16.7 para encontrar la probabilidad de que el fabricante tenga que cubrir una garantía.

16.8-1. Lea el artículo de referencia que describe el estudio de IO que se resume en el recuadro de aplicación que se presentó en la sección 16.8. Describa de manera breve la forma en que se aplicó una cadena de Markov de tiempo continuo a este estudio. Después, elabore una lista de los beneficios financieros y de otro tipo que arrojó dicho estudio.

16.8-2. Reconsidere el ejemplo que se presentó al final de la sección 16.8. Suponga que ahora se agrega al taller una tercera máquina, idéntica a las dos primeras. La persona de mantenimiento debe atender todas las máquinas.

a) Desarrolle un *diagrama de tasas* de esta cadena de Markov.
b) Construya las *ecuaciones de estado estable*.
c) Resuelva estas ecuaciones para obtener las *probabilidades de estado estable*.

16.8-3. El estado de una cadena de Markov de tiempo continuo está definido como el número de trabajos que hay en el momento actual en cierto centro de trabajo, donde se permite un máximo de dos trabajos, los cuales llegan individualmente. Siempre que hay menos de tres trabajos, el tiempo que transcurre hasta la siguiente llegada tiene una distribución exponencial con media de dos días. Los trabajos se procesan uno a la vez y dejan el centro de inmediato. Los tiempos de procesado tienen una distribución exponencial con media de un día.

***a*)** Construya el *diagrama de tasas* de esta cadena de Markov.
***b*)** Escriba las *ecuaciones de estado estable*.
***c*)** Resuelva estas ecuaciones para obtener las *probabilidades de estado estable.*

CAPÍTULO 17

Teoría de colas

Las *colas* (líneas de espera) son parte de la vida diaria. Todos esperamos en colas para comprar un boleto para el cine, hacer un depósito en el banco, pagar en el supermercado, enviar un paquete por correo, obtener comida en la cafetería, subir a un juego en la feria, etc. Nos hemos acostumbrado a una considerable cantidad de esperas, pero todavía nos molesta cuando éstas son demasiado largas.

Sin embargo, tener que esperar no sólo es una molestia personal. El tiempo que la población de un país pierde al esperar en las colas es un factor importante tanto de la calidad de vida como de la eficiencia de su economía.

También ocurren grandes ineficiencias debido a otros tipos de espera que no son personas en una cola. Por ejemplo, cuando las *máquinas* esperan ser reparadas pueden provocarse pérdidas de producción. Los *vehículos* (incluso barcos y camiones) que deben esperar su descarga pueden retrasar envíos subsecuentes. Los *aviones* que esperan despegar o aterrizar pueden desorganizar la programación posterior de vuelos. Los retrasos de las transmisiones de *telecomunicaciones* por saturación de líneas pueden causar fallas inesperadas en los datos. Cuando los *trabajos de manufactura* esperan su proceso se puede perturbar el proceso de producción. El retraso de los *trabajos de servicio* respecto de su fecha de entrega es una causa de pérdida de negocios futuros.

La *teoría de colas* es el estudio de la espera en las distintas modalidades. Utiliza los *modelos de colas* para representar los tipos de *sistemas de líneas de espera* (sistemas que involucran colas de algún tipo) que surgen en la práctica. Las fórmulas de cada modelo indican cuál debe ser el desempeño del sistema correspondiente y señalan la cantidad promedio de espera que ocurrirá en diversas circunstancias.

Por lo tanto, estos modelos de líneas de espera son muy útiles para determinar cómo operar un sistema de colas de la manera más eficaz. Proporcionar demasiada capacidad de servicio para operar el sistema implica costos excesivos; pero si no se cuenta con suficiente capacidad de servicio surgen esperas excesivas con todas sus desafortunadas consecuencias. Los modelos permiten encontrar un balance adecuado entre el costo de servicio y la cantidad de espera.

Después de una exposición general, en este capítulo se presenta la mayoría de los modelos de líneas de espera elementales y sus resultados básicos. En la sección 17.10 se estudia cómo puede usarse la información que proporciona la teoría de colas para diseñar sistemas que minimicen el costo total de servicio y espera. Después, en el capítulo 26 (en el sitio en internet de este libro) se profundiza en la aplicación de la teoría de colas en este sentido.

17.1 EJEMPLO PROTOTIPO

La sala de urgencias del HOSPITAL GENERAL proporciona cuidados médicos rápidos a los casos de emergencia que llegan en ambulancia o vehículos particulares. En todo momento se cuenta con un médico de guardia. No obstante, debido a la creciente tendencia a usar estas instalaciones para casos de urgencia en lugar de ir a una clínica privada, cada año el hospital experimenta un

aumento continuo del número de pacientes que llegan a la sala de emergencias. Como resultado, es bastante común que los pacientes que llegan durante las horas pico (temprano en la tarde) tengan que esperar turno para recibir el tratamiento del médico. Por ello, se ha presentado una propuesta para asignar un segundo médico a esta sala durante esas horas pico, para que se puedan atender dos casos de emergencia al mismo tiempo. Se ha pedido al ingeniero administrador del hospital que estudie esta opción.

El ingeniero comenzó por reunir los datos históricos pertinentes y hacer una proyección de ellos al siguiente año. Reconoció que la sala de urgencias es un sistema de líneas de espera y aplicó varios modelos de teoría de colas para predecir las características de la espera en el sistema con uno y dos médicos, como se verá más adelante en este capítulo (vea las tablas 17.2 y 17.3).

17.2 ESTRUCTURA BÁSICA DE LOS MODELOS DE COLAS

Proceso básico de colas

El proceso básico supuesto por la mayoría de los modelos de colas es el siguiente. Los *clientes* que requieren un servicio se generan en el tiempo en una *fuente de entrada*. Luego, entran al *sistema* y se unen a una *cola*. En determinado momento se selecciona un miembro de la cola para proporcionarle el servicio mediante alguna regla conocida como *disciplina de la cola*. Se lleva a cabo el servicio que el cliente requiere mediante un *mecanismo de servicio*, y después el cliente sale del sistema de colas. En la figura 17.1 se describe este proceso.

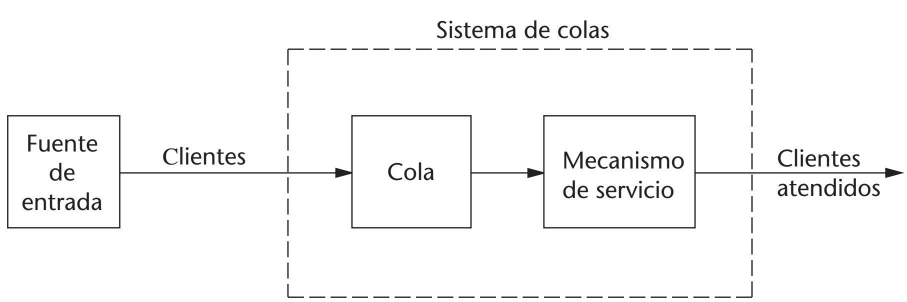
Se pueden hacer muchos supuestos sobre los distintos elementos del proceso de colas que se analizarán a continuación.

Fuente de entrada (población potencial)

Una característica de la fuente de entrada es su tamaño. El *tamaño* es el número total de clientes que pueden requerir servicio en determinado momento, es decir, el número total de clientes potenciales. Esta población a partir de la cual surgen las unidades que llegan se conoce como **población de entrada**. Puede suponerse que el tamaño es *infinito* o *finito* (de modo que también se dice que la fuente de entrada es *ilimitada* o *limitada*). Debido a que los cálculos son mucho más sencillos en el caso del tamaño infinito, este supuesto se hace a menudo aun cuando el tamaño real sea un número fijo relativamente grande, y debe tomarse como un supuesto implícito en cualquier modelo en el que no se establezca otra cosa. Desde una perspectiva analítica, el caso finito es más complejo puesto que el número de clientes que conforman la cola afecta al número potencial de clientes fuera del sistema en cualquier momento; pero debe hacerse este supuesto de finitud si la tasa a la que la fuente de entrada genera clientes nuevos es afectada en forma significativa por el número de clientes existentes en el sistema de líneas de espera.

También se debe especificar el patrón estadístico mediante el cual se generan los clientes en el tiempo. El supuesto normal es que se generan de acuerdo con un *proceso Poisson*; es decir, el número de clientes que llegan hasta un momento específico tiene una distribución de Poisson. Como se analizará en la sección 17.4, este caso corresponde a aquel cuyas llegadas al sistema ocurren de manera aleatoria pero con cierta tasa media fija y sin que importe cuántos clientes están ya ahí (por lo que el *tamaño* de la fuente de entrada es *infinito*). Un supuesto equivalente es que la distribución

FIGURA 17.1
Proceso básico de colas.



de probabilidad del tiempo que transcurre entre dos llegadas consecutivas es *exponencial*. (En la sección 17.4 se describen las propiedades de esta distribución.) Se hace referencia al tiempo que transcurre entre dos llegadas consecutivas como **tiempo entre llegadas**.

También debe especificarse cualquier otro supuesto no usual sobre el comportamiento de los clientes. Un ejemplo sería cuando se pierde un cliente porque *desiste o se rehúsa* a entrar al sistema porque la cola es demasiado larga.

Cola

La cola es donde los clientes esperan *antes* de recibir el servicio. Una cola se caracteriza por el número máximo permisible de clientes que puede admitir. Las colas pueden ser *finitas* o *infinitas*, según si dicho número es finito o infinito. El supuesto de una *cola infinita* es el estándar de la mayoría de los modelos, incluso en situaciones en las que en realidad existe una cota superior (relativamente grande) sobre el número permitido de clientes, puesto que manejar una cota así puede ser un factor que complique el análisis. En los sistemas de colas en los que la cota superior es tan pequeña que se llega a ella con cierta frecuencia, es necesario suponer una *cola finita*.

Disciplina de la cola

La disciplina de la cola se refiere al orden en el que sus miembros se seleccionan para recibir el servicio. Por ejemplo, puede ser: primero en entrar, primero en salir; aleatoria; de acuerdo con algún procedimiento de prioridad o con algún otro orden. En los modelos de colas se supone como normal a la disciplina de primero en entrar, primero en salir, a menos que se establezca de otra manera.

Mecanismo de servicio

El mecanismo de servicio consiste en una o más *estaciones de servicio*, cada una de ellas con uno o más *canales de servicio paralelos*, llamados **servidores**. Si existe más de una estación de servicio, el cliente puede recibirlo de una secuencia de ellas (*canales de servicio en serie*). En una estación dada, el cliente entra en uno de estos canales y el servidor le presta el servicio completo. Los modelos de colas deben especificar el arreglo de las estaciones y el número de servidores (canales paralelos) en cada una de ellas. Los modelos más elementales suponen una estación, ya sea con un servidor o con un número finito de servidores.

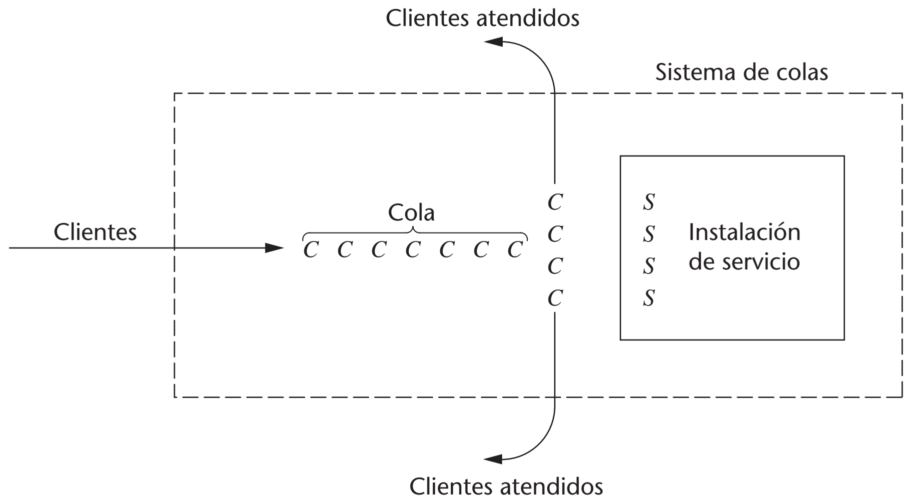
El tiempo que transcurre desde el inicio del servicio para un cliente hasta su terminación en una estación se llama **tiempo de servicio** (o *duración del servicio*). Un modelo de un sistema de colas determinado debe especificar la distribución de probabilidad de los tiempos de servicio de cada servidor (y tal vez de los distintos tipos de clientes), aunque es común suponer la *misma* distribución para todos los servidores (todos los modelos en este capítulo se basan en este supuesto). La distribución del tiempo de servicio que más se usa en la práctica (por ser más manejable que cualquier otra) es la distribución *exponencial* que se presenta en la sección 17.4, por lo que casi todos los modelos de este capítulo serán de este tipo. Otras distribuciones de tiempos de servicio importantes son la distribución *degenerada* (tiempos de servicio constantes) y la distribución *Erlang* (gamma) que se ilustran en los modelos de la sección 17.7.

El proceso de colas elemental

Como ya se ha señalado, la teoría de colas se aplica a muchos tipos diferentes de situaciones. El tipo que prevalece es el siguiente: una sola línea de espera (que a veces puede estar vacía) se forma frente a una estación de servicio, dentro de la cual se encuentra uno o más servidores. Cada cliente generado por una fuente de entrada recibe el servicio de uno de los servidores, quizá después de esperar un poco en la cola (línea de espera). En la figura 17.2 se presenta un esquema del sistema de colas del que se habla.

Observe que el proceso que se ilustra en el ejemplo de la sección 17.1 es de este tipo. La fuente de entrada genera clientes en la forma de casos urgentes que requieren cuidado médico. La sala de urgencias es la instalación de servicio y los médicos son los servidores.

Un servidor no tiene que ser un solo individuo; puede ser un grupo de personas, por ejemplo, una cuadrilla de reparación que combina fuerzas para realizar, de manera simultánea, el servicio que solicita el cliente. Aún más, los servidores ni siquiera tienen que ser personas. En muchos

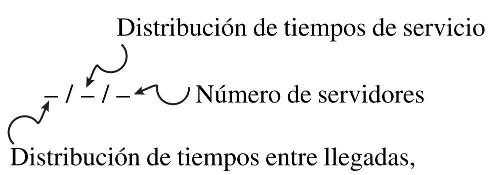


■ **FIGURA 17.2**
Sistema de colas elemental (cada cliente se indica con una *C* y cada servidor con una *S*).

casos puede ser una máquina, un vehículo, un dispositivo electrónico, etc. En esta misma línea de ideas, los clientes que conforman la cola no tienen que ser personas. Por ejemplo, pueden ser unidades que esperan ser procesadas en cierto tipo de máquina, o automóviles que deben pasar por una caseta de cobro.

En realidad, no es necesario que se forme una línea de espera física delante de una estructura material que constituye la estación de servicio. Los miembros de la cola pueden estar dispersos en un área mientras esperan que el servidor venga a ellos, como las máquinas que esperan reparación. El servidor o grupo de servidores asignados a un área constituyen la estación de servicio de esa área. De todas maneras, la teoría de colas proporciona, entre otros, un número promedio de clientes en espera —el tiempo promedio de espera—, puesto que es irrelevante si los clientes esperan en grupo o no. El único requisito esencial para poder aplicar la teoría de colas es que los cambios en el número de clientes que esperan un servicio ocurran como si prevaleciera la situación física que se describe en la figura 17.2 (o una contraparte válida).

Con excepción de la sección 17.9, todos los modelos de colas que se presentan en este capítulo son del tipo elemental que se esquematiza en la figura 17.2. Muchos de ellos se basan en el supuesto de que todos los *tiempos entre llegadas* y todos los *tiempos de servicio* son independientes e idénticamente distribuidos. Por convención, estos modelos se etiquetan de la siguiente manera:



donde: M = distribución exponencial (markoviana), como se describe en la sección 17.4,

D = distribución degenerada (tiempos constantes), como se expone en la sección 17.7,

E_k = distribución Erlang (parámetro de forma = k), como se describe en la sección 17.7,

G = distribución general (permite cualquier distribución arbitraria),[1] como se presenta en la sección 17.7.

[1] Al referirse a los tiempos entre llegadas, se sustituyó, de manera convencional, el símbolo G por GI = distribución general independiente.

Por ejemplo, el modelo *M/M/s* que se presenta en la sección 17.6 supone que tanto los tiempos entre llegadas como los de servicio tienen distribución exponencial y que el número de servidores es *s* (cualquier entero positivo). El modelo *M/G/*1 que se estudia en la sección 17.7 supone que los tiempos entre llegadas siguen una distribución exponencial pero no pone restricciones a la distribución de los tiempos de servicio, mientras que el número de servidores está restringido a sólo 1. En la sección 17.7 se presentan algunos otros modelos que se ajustan a este esquema de etiquetas.

Terminología y notación

A menos que se establezca otra cosa, se utilizará la siguiente terminología estándar:

Estado del sistema = número de clientes en el sistema.

Longitud de la cola = número de clientes que esperan servicio.

= estado del sistema *menos* número de clientes a quienes se les da el servicio.

$N(t)$ = número de clientes en el sistema de colas en el tiempo t ($t \geq 0$).

$P_n(t)$ = probabilidad de que exactamente n clientes estén en el sistema en el tiempo t, dado el número en el tiempo 0.

s = número de servidores (canales de servicio en paralelo) en el sistema de colas.

λ_n = tasa media de llegadas (número esperado de llegadas por unidad de tiempo) de nuevos clientes cuando hay n clientes en el sistema.

μ_n = tasa media de servicio en todo el sistema (número esperado de clientes que completan su servicio por unidad de tiempo) cuando hay n clientes en el sistema. *Nota*: μ_n representa la tasa *combinada* a la que todos los servidores *ocupados* (aquellos que están sirviendo a un cliente) logran terminar sus servicios.

λ, μ, ρ = vea el párrafo siguiente.

Cuando λ_n es constante para toda n, esta constante se denota por λ. Cuando la tasa media de servicio *por servidor ocupado* es constante para toda $n \geq 1$, esta constante se denota por μ. (En este caso, $\mu_n = s\mu$ cuando $n \geq s$, es decir, cuando los s servidores están ocupados.) En estas circunstancias, $1/\lambda$ y $1/\mu$ es el *tiempo esperado entre llegadas* y el *tiempo esperado de servicio*, respectivamente. Asimismo, $\rho = \lambda/(s\mu)$ es el **factor de utilización** de la instalación de servicio, es decir, la fracción esperada de tiempo que los servidores individuales están ocupados, puesto que $\lambda/(s\mu)$ representa la fracción de la capacidad de servicio del sistema ($s\mu$) que *utilizan* en promedio los clientes que llegan (λ).

También se requiere cierta notación para describir los resultados de *estado estable*. Cuando un sistema de colas apenas inicia su operación, el estado del sistema (el número de clientes que esperan en el sistema) se encuentra bastante afectado por el estado inicial y el tiempo que ha pasado desde el inicio. Se dice entonces que el sistema se encuentra en **condición transitoria**. Sin embargo, una vez que ha pasado suficiente tiempo, el estado del sistema se vuelve, en esencia, independiente del estado inicial y del tiempo transcurrido (excepto en circunstancias no usuales).[2] En este contexto, se puede decir que el sistema ha alcanzado su **condición de estado estable**, en la que la distribución de probabilidad del estado del sistema se conserva (la distribución *estacionaria* o de *estado estable*) a través del tiempo. La teoría de colas tiende a dedicar su análisis a la condición de estado estable, en parte porque el caso transitorio es analíticamente más difícil. (Existen algunos resultados transitorios pero en general están más allá del alcance de este libro.) La notación siguiente supone que el sistema se encuentra en la *condición de estado estable*:

P_n = probabilidad de que haya exactamente n clientes en el sistema.

L = número esperado de clientes en el sistema $= \sum_{n=0}^{\infty} nP_n$.

[2] Cuando λ y μ están definidos, estas circunstancias poco usuales se refieren a que $\rho \geq 1$, en cuyo caso el estado del sistema tiende a crecer en forma continua conforme pasa el tiempo.

L_q = longitud esperada de la cola (excluye los clientes que están en servicio)

$$= \sum_{n=s}^{\infty} (n - s)P_n.$$

$\mathcal{W}$ = tiempo de espera en el sistema (incluye tiempo de servicio) para cada cliente.

$W = E(\mathcal{W})$.

$\mathcal{W}_q$ = tiempo de espera en la cola (excluye tiempo de servicio) para cada cliente.

$W_q = E(\mathcal{W}_q)$.

Relaciones entre *L*, *W*, L_q, y W_q

Suponga que λ_n es una constante λ para toda n. Se ha demostrado que en un proceso de colas en estado estable,

$$L = \lambda W.$$

(Dado que John D. C. Little proporcionó la primera demostración rigurosa, a veces se le da el nombre de **fórmula de Little**.) Además, la misma demostración prueba que

$$L_q = \lambda W_q.$$

Si las λ_n no son iguales, entonces λ se puede sustituir en estas ecuaciones por $\bar{\lambda}$, la tasa *promedio* entre llegadas a largo plazo. (Más adelante se verá cómo se puede determinar $\bar{\lambda}$ en algunos casos básicos.)

Ahora suponga que el tiempo medio de servicio es una constante $1/\mu$, para toda $n \geq 1$. Se tiene entonces que

$$W = W_q + \frac{1}{\mu}.$$

Estas relaciones son en extremo importantes, puesto que permiten determinar las cuatro cantidades fundamentales: L, W, L_q y W_q en cuanto se encuentra analíticamente el valor de una de ellas. Esta situación es afortunada, ya que suele ser mucho más fácil determinar una de ellas que las otras al resolver un modelo de colas a partir de los principios básicos.

17.3 EJEMPLOS DE SISTEMAS DE COLAS REALES

Puede parecer que la descripción de los sistemas de colas de la sección 17.2 es algo abstracta y que sólo es aplicable en situaciones prácticas bastante especiales. Por el contrario, los sistemas de colas se aplican con sorprendente frecuencia en una amplia variedad de contextos. Para ampliar el horizonte sobre sus aplicaciones, se mencionarán brevemente varios ejemplos reales de sistemas de colas que pertenecen a varias categorías generales. Después se describirán sistemas de colas en algunas compañías prominentes (y en una ciudad) y los estudios premiados que se llevaron a cabo para diseñar estos sistemas.

Algunas clases de sistemas de colas

Una clase importante de sistemas de colas que se encuentra en la vida diaria es el **sistema de servicio comercial**, en donde los clientes externos reciben un servicio de una organización comercial. Muchos de estos sistemas incluyen un servicio de persona a persona en un local fijo, como una peluquería (los peluqueros son los servidores), el servicio de una cajera de banco, las cajas de cobro de un supermercado y una cola en una cafetería (canales de servicio en serie). Sin embargo, muchos otros sistemas son de un tipo diferente, como la reparación de aparatos domésticos (el servidor va hacia el cliente), una máquina de monedas (el servidor es una máquina) y una gasolinera (los clientes son automóviles).

Otra clase importante es la de **sistemas de servicio de transporte**. En algunos de estos sistemas los vehículos son los clientes, como los automóviles que esperan para pasar por una caseta de cobro o un semáforo (el servidor), un camión de carga o un barco que esperan que una cuadrilla

les dé el servicio de carga o descarga y un avión que espera aterrizar o despegar en una pista (el servidor). (Un estacionamiento es un ejemplo poco usual de este tipo, en el que los automóviles son los clientes y los espacios son los servidores, pero no existe una cola porque si un estacionamiento está lleno, los clientes se van a otro.) En otros casos, los vehículos son los servidores, como los taxis, los camiones de bomberos y los elevadores.

En años recientes, la teoría de colas se ha aplicado más a los **sistemas de servicio interno** donde los clientes que reciben el servicio son *personal interno* o parte de la organización. Los ejemplos incluyen sistemas de manejo de materiales, en donde las unidades de manejo de materiales (los servidores) mueven cargas (los clientes); sistemas de mantenimiento, en los cuales las brigadas de mantenimiento (los servidores) reparan máquinas (los clientes) y puestos de inspección en los que los inspectores de control de calidad (los servidores) inspeccionan artículos (los clientes). Las instalaciones para empleados y los departamentos que les prestan servicio también entran en esta categoría. Además, las máquinas se pueden ver como servidores cuyos clientes son los trabajos que están procesando. Un ejemplo relacionado muy importante es un centro de cómputo en el que la computadora se puede ver como el servidor.

Existe un reconocimiento creciente de que la teoría de colas también se puede aplicar a **sistemas de servicio social**. Por ejemplo, un sistema judicial es una red de colas, donde las cortes son las instalaciones de servicio, los jueces (o los jurados) son los servidores y los casos que esperan el proceso son los clientes. Un sistema legislativo es una red de colas similar, en el cual los clientes son los asuntos que el congreso va a tratar. Algunos sistemas de salud pública son sistemas de colas. En la sección 17.1 se vio un ejemplo (la sala de urgencias de un hospital), pero también las ambulancias, las máquinas de rayos X y las camas del hospital pueden actuar como servidores en sus propios sistemas. En forma parecida, las familias en espera de viviendas de interés social u otros servicios pueden ser clientes de un sistema de colas.

Aun cuando éstas son cuatro clases amplias de sistemas de colas, la lista todavía no se agota. En realidad, la teoría de colas comenzó a principios de siglo con aplicaciones a ingeniería telefónica (el fundador de la teoría de colas, A. K. Erlang, era empleado de la Danish Telephone Company, en Copenhague), y la ingeniería telefónica constituye todavía una importante aplicación. Lo que es más, cada individuo tiene sus propias líneas de espera personales: tareas, libros que leer, etc. Estos ejemplos son suficientes para sugerir que los sistemas de colas sin duda se presentan con toda frecuencia en muchas áreas de la sociedad.

Algunas aplicaciones de teoría de colas ganadoras de premios

El prestigioso premio *Franz Edelman Awards for Management Science Achievement* es otorgado cada año por el Institute of Operations Research and Management Sciences (INFORMS) a la mejor aplicación de IO del año. Un buen número de estos premios se ha otorgado por aplicaciones innovadoras de la teoría de colas al diseño de sistema de colas.

Dos de dichas aplicaciones de la teoría de colas ganadoras de premios se describen en las viñetas de aplicación más adelante en este capítulo (secciones 17.6 y 17.9). Las referencias seleccionadas al final del capítulo también incluyen una muestra de artículos que describen otras aplicaciones ganadoras de premios. (En el sitio en internet de este libro se proporciona un enlace hacia todos estos artículos, incluyendo las viñetas de aplicación.) A continuación se describe, en forma breve, algunas de dichas aplicaciones de la teoría de colas.

Como se mencionó en la referencia seleccionada A1, uno de los primeros ganadores del premio en la competencia Edelman fue *Xerox Corporation*. La compañía acababa de introducir un sistema de duplicado nuevo que sería en especial valioso para quienes lo compraran. Por lo tanto, los clientes demandaban que los representantes técnicos de Xerox redujeran los tiempos de espera de la reparación de máquinas. Un equipo de IO aplicó la teoría de colas para estudiar la mejor manera de cumplir con este nuevo requerimiento de servicio. El resultado fue reemplazar la asignación previa de territorios a un técnico por territorios más grandes asignados a tres personas. Este cambio tuvo un efecto drástico en la reducción del tiempo de espera promedio de los clientes y en el incremento de más de 50% del empleo de los técnicos. (El capítulo 11 de la referencia seleccionada 9 presenta el estudio de un caso que se basa en esta aplicación de la teoría de colas por parte de Xerox Corporation.)

L. L. Bean, Inc., la empresa más grande de telemercadeo y ventas por catálogo, realizó un estudio ganador de premio cuyo apoyo primordial fue la teoría de colas para determinar cómo

asignar sus recursos de telecomunicaciones, tema que se describe en la referencia seleccionada A4. Las llamadas telefónicas que llegan al centro para hacer pedidos son los clientes de un gran sistema de colas, mientras que los agentes son los servidores. Las preguntas clave que se hicieron durante el estudio fueron las siguientes:

1. ¿Con cuántas líneas troncales de teléfono se debe contar para atender las llamadas que entran al centro?
2. ¿Cuántos agentes de ventas por teléfono deben programarse en diferentes periodos?
3. ¿Cuántas posiciones para llamadas en espera debe haber para que los clientes esperen a un agente? (Observe que el número limitado de llamadas en espera ocasiona que el sistema tenga una cola finita.)

Por cada combinación interesante de estas tres cantidades, los modelos de colas proporcionan una medida de desempeño del sistema. Dadas estas medidas, el equipo de IO evaluó con cuidado el costo de las ventas perdidas debido a que algunos clientes escucharan tono de ocupado o de llamada en espera por mucho tiempo. Cuando se agregó el costo de los recursos de telemercadeo, el equipo pudo encontrar la combinación de las tres cantidades que minimiza el costo total esperado. El resultado fue un ahorro anual de costos de 9 a 10 millones de dólares.

Otro ganador del primer premio del concurso Edelman fue AT&T por un estudio que combinaba el uso de la teoría de colas y la simulación (tema del capítulo 20). Como se describió en la referencia seleccionada A2, los modelos de colas se aplicaron tanto a la red de telecomunicaciones de AT&T como al entorno de centros de llamadas de los clientes de la compañía que tenían estos centros. El objetivo del estudio fue desarrollar un sistema amigable de PC que los clientes pudieran usar para que la empresa los guiara en el proceso de diseñar o rediseñar sus centros de llamadas. Como éstos constituyen una de las industrias de más rápido crecimiento en Estados Unidos, para cuando se escribió este artículo, los clientes de negocios de AT&T habían usado este sistema más de 2 000 veces. El resultado fue una ganancia anual de más de 750 millones de dólares generada por estos clientes.

Hewlett-Packard (HP) es una manufacturera de equipo electrónico multinacional líder en su ramo. En 1993 la compañía instaló un sistema de línea de ensamble mecanizado para fabricar impresoras de inyección de tinta en su planta de Vancouver, Washington, para satisfacer la explosiva demanda de este tipo de impresoras. Pronto se hizo evidente que el sistema instalado no sería tan rápido ni tan confiable como para satisfacer las metas de producción de la compañía. En consecuencia, se formó un equipo conjunto de científicos de la administración de HP y del Massachusetts Institute of Technology (MIT) para estudiar cómo rediseñar el sistema para mejorar su desempeño.

Como se describe en la referencia seleccionada A3 sobre este estudio ganador de premios, el equipo de HP/MIT se percató con rapidez de que el sistema de línea de ensamble se podría modelar como un tipo especial de sistema de colas donde los clientes (las impresoras que debían ensamblarse) pasarían a través de una serie de servidores (operaciones de ensamblaje) en una secuencia fija. Un modelo de colas especial para este tipo de sistema proporcionó con rapidez los resultados analíticos que se necesitaban para determinar el modo en que el sistema debería rediseñarse para lograr la capacidad que se requería de la forma más económica. Los cambios incluyeron la incorporación de espacios de almacenamiento en puntos estratégicos para mantener de mejor manera el flujo de trabajo hacia las estaciones subsecuentes y para contrarrestar el efecto de la descompostura de las máquinas. El nuevo diseño incrementó alrededor de 50% la productividad y produjo aumentos en las ganancias de aproximadamente 280 millones de dólares por ventas de impresoras así como utilidades adicionales por productos accesorios. Esta aplicación innovadora del modelo de colas especial también le dio a HP un nuevo método para crear diseños de sistemas rápidos y eficaces en otras áreas de la compañía.

17.4 PAPEL DE LA DISTRIBUCIÓN EXPONENCIAL

En gran medida, las características operativas de los sistemas de colas están determinadas por dos propiedades estadísticas, a saber, la distribución de probabilidad de los *tiempos entre llegadas* (vea "Fuente de entrada" en la sección 17.2) y la distribución de probabilidad de los *tiempos de servicio* (vea "Mecanismo de servicio" en la sección 17.2). En los sistemas de colas reales, estas distribucio-

nes pueden tomar casi cualquier forma. (La única restricción es que no pueden presentarse valores negativos.) Sin embargo, para formular un *modelo* de teoría de colas como una representación del sistema real, es necesario especificar la forma supuesta de cada una de estas distribuciones. Para que sea útil, la forma supuesta debe ser lo *suficientemente realista* como para que el modelo proporcione *predicciones razonables*, pero al mismo tiempo debe ser lo *suficientemente sencilla* para que sea *matemáticamente manejable*. Con estas consideraciones en mente, la distribución de probabilidad más importante en la teoría de colas es la *distribución exponencial*.

Suponga que una variable aleatoria T representa ya sea los tiempos entre llegadas o los tiempos de servicio. (Se hace referencia a los hechos que marcan el final de estos tiempos, de llegadas o de terminación de un servicio, como *eventos*.) Se dice que esta variable aleatoria tiene una *distribución exponencial con parámetro* α si su función de densidad de probabilidad es

$$f_T(t) = \begin{cases} \alpha e^{-\alpha t} & \text{para } t \geq 0 \\ 0 & \text{para } t < 0, \end{cases}$$

como se muestra en la figura 17.3. En este caso, las probabilidades acumuladas son

$$\begin{aligned} P\{T \leq t\} &= 1 - e^{-\alpha t} \\ P\{T > t\} &= e^{-\alpha t} \end{aligned} \qquad (t \geq 0),$$

y el valor esperado y la variancia de T son

$$E(T) = \frac{1}{\alpha},$$

$$\text{var}(T) = \frac{1}{\alpha^2}.$$

¿Cuáles son las implicaciones para el modelo de colas si se supone que T tiene una distribución exponencial? Para explorar esta pregunta se examinarán seis propiedades de la distribución exponencial.

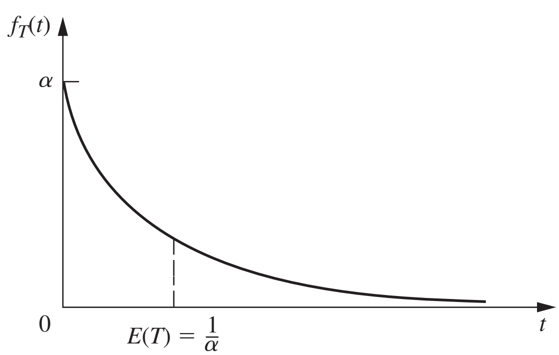
Propiedad 1: $f_T(t)$ es una función de t estrictamente *decreciente* de t $(t \geq 0)$.

Una consecuencia de la propiedad 1 es que

$$P\{0 \leq T \leq \Delta t\} > P\{t \leq T \leq t + \Delta t\}$$

para cualesquiera valores estrictamente positivos de Δt y t. [Ésta es una consecuencia del hecho de que estas probabilidades son el área bajo la curva de $f_T(t)$ en el intervalo indicado de longitud Δt, y la altura promedio de la curva es menor para la segunda probabilidad que para la primera.] Por lo tanto, no sólo es posible sino también bastante probable que T tome un valor pequeño cercano a cero. En realidad,

$$P\left\{0 \leq T \leq \frac{1}{2}\,\frac{1}{\alpha}\right\} = 0.393$$



■ **FIGURA 17.3**
Función de densidad de probabilidad de la distribución exponencial.

mientras que

$$P\left\{\frac{1}{2}\frac{1}{\alpha} \leq T \leq \frac{3}{2}\frac{1}{\alpha}\right\} = 0.383,$$

de manera que es más probable que el valor que tome T sea "pequeño" [esto es, menor que la mitad de $E(T)$] que "cercano" a su valor esperado [es decir, no más alejado que la mitad de $E(T)$], aun cuando el segundo intervalo tiene el doble de amplitud que el primero.

¿En realidad, ésta es una propiedad razonable de T en un modelo de colas? Si T representa los *tiempos de servicio*, la respuesta depende de la naturaleza general del servicio en cuestión, como se verá a continuación.

Si el servicio que se requiere es en esencia idéntico para cada cliente y el servidor realiza siempre la misma secuencia de operaciones, entonces los tiempos de servicio reales tienden a ser cercanos al tiempo esperado de servicio. Pueden ocurrir pequeñas desviaciones de la media, pero por lo general se deben a variaciones menores en la eficiencia del servidor. Un tiempo de servicio tan pequeño que quede muy por debajo de la media es en realidad imposible, puesto que se necesita cierta cantidad mínima de tiempo para realizar las operaciones de servicio que se requieren, aunque el servidor trabaje a la mayor velocidad. Es claro que la distribución exponencial no proporciona una aproximación cercana a la distribución de tiempos de servicio en este tipo de situación.

Por otro lado, considere el tipo de situación en la que las tareas específicas que tiene que realizar el servidor difieren de un cliente a otro. La naturaleza general del servicio puede ser la misma, pero la cantidad y tipo específico de servicio difieren. Por ejemplo, éste es el caso en el problema de la sala de emergencia del Hospital General que se presentó en la sección 17.1. El médico se enfrenta a una gran variedad de problemas de su profesión. En la mayor parte de los casos puede proporcionar el tratamiento que se requiere con bastante rapidez pero, en ocasiones, el paciente necesita un cuidado más especializado. De igual manera, los supervisores de bancos y supermercados son servidores de este tipo general, en donde el servicio que prestan suele ser breve, pero en ocasiones se extiende. Parece posible una distribución exponencial de los tiempos de servicio en este tipo de situación.

Si T representa los *tiempos entre llegadas*, la propiedad 1 descarta las situaciones en las que los clientes que llegan al sistema tienden a posponer su entrada si ven que otro cliente entra antes que ellos. Por otro lado, es totalmente congruente con el fenómeno común de las llegadas "aleatorias" que se describe con las propiedades subsecuentes. Así, cuando se grafican los tiempos entre llegadas contra el tiempo, a veces tienen la apariencia de estar aglomerados con grandes separaciones entre cada aglomeración, debido a la gran probabilidad de que los tiempos entre llegadas sean pequeños y la poca probabilidad de que ocurran tiempos entre llegadas grandes, pero un patrón tan irregular es exactamente parte de la verdadera aleatoriedad.

Propiedad 2: Falta de memoria.

Esta propiedad se puede expresar en forma matemática como

$$P\{T > t + \Delta t \,|\, T > \Delta t\} = P\{T > t\}$$

para cualesquiera cantidades positivas de t e Δt. En otras palabras, la distribución de probabilidad del tiempo que falta hasta que ocurra el evento (llegada o terminación de servicio) siempre es la misma, sin importar cuánto tiempo (Δt) haya pasado. En efecto, el proceso "olvida" su historia. Este sorprendente fenómeno ocurre con la distribución exponencial debido a que

$$\begin{aligned} P\{T > t + \Delta t \,|\, T > \Delta t\} &= \frac{P\{T > \Delta t, T > t + \Delta t\}}{P\{T > \Delta t\}} \\ &= \frac{P\{T > t + \Delta t\}}{P\{T > \Delta t\}} \\ &= \frac{e^{-\alpha(t+\Delta t)}}{e^{-\alpha \Delta t}} \\ &= e^{-\alpha t} \\ &= P\{T > t\}. \end{aligned}$$

En el caso de los *tiempos entre llegadas*, esta propiedad describe la situación común en donde el tiempo que transcurre hasta la siguiente llegada está totalmente influenciado por el momento en que ocurrió la última llegada. En el caso de los *tiempos de servicio*, esta propiedad es más difícil de interpretar. No debe esperarse que se cumpla cuando el servidor tiene que realizar la misma secuencia fija de operaciones para cada cliente, porque entonces un servicio largo y lento debe implicar que tal vez queda muy poco por hacer. Sin embargo, en la clase de situación en la que las operaciones de servicio que se requieren difieren entre los clientes, la afirmación matemática de la propiedad es bastante realista. En este caso, si ha pasado un tiempo de servicio considerable, la única implicación puede ser que este cliente en particular requiera un servicio más extenso que los demás.

Propiedad 3: El *mínimo* de diversas variables aleatorias exponenciales independientes tiene una distribución exponencial.

Para establecer matemáticamente esta propiedad, sean $T_1, T_2, \ldots, T_n$ variables aleatorias exponenciales *independientes* con parámetros $\alpha_1, \alpha_2, \ldots, \alpha_n$, respectivamente. También sea U la variable aleatoria cuyo valor es igual al *mínimo* de los valores que toman $T_1, T_2, \ldots, T_n$; es decir,

$$U = \text{mín}\ \{T_1, T_2, \ldots, T_n\}.$$

Así, si T_i representa el tiempo que pasa hasta que ocurre un tipo especial de evento, entonces U representa el tiempo que pasa hasta que ocurre el *primero* de los n eventos diferentes. Ahora, observe que para cualquier $t \geq 0$,

$$\begin{aligned} P\{U > t\} &= P\{T_1 > t, T_2 > t, \ldots, T_n > t\} \\ &= P\{T_1 > t\}P\{T_2 > t\} \cdots P\{T_n > t\} \\ &= e^{-\alpha_1 t}e^{-\alpha_2 t} \cdots e^{-\alpha_n t} \\ &= \exp\left(-\sum_{i=1}^{n} \alpha_i t\right), \end{aligned}$$

de manera que, sin duda, U tiene distribución exponencial con parámetro

$$\alpha = \sum_{i=1}^{n} \alpha_i.$$

Esta propiedad tiene algunas implicaciones para los tiempos entre llegadas en los modelos de colas. En particular, suponga que existen varios (n) tipos *diferentes* de clientes, pero que los tiempos entre llegadas de *cada* tipo (tipo i) tienen distribución exponencial con parámetro α_i ($i = 1, 2, \ldots, n$). De acuerdo con la propiedad 2, el tiempo *que falta* a partir de un instante específico hasta la llegada del siguiente cliente del tipo i tendrá esta misma distribución. Por ello, sea T_i este tiempo restante medido a partir del instante en que llega un cliente de *cualquier* tipo. La propiedad 3 dice entonces que U, el tiempo entre llegadas del sistema de colas completo, tiene distribución exponencial con parámetro α definido por la última ecuación. Como resultado, se puede hacer caso omiso de la distinción entre los clientes y seguir teniendo tiempos entre llegadas exponenciales en el modelo de colas.

Sin embargo, estas implicaciones son todavía más importantes para los *tiempos de servicio* en los modelos de colas que tienen más de un servidor, de lo que son para los tiempos entre llegadas. Por ejemplo, considere la situación en la que todos los servidores tienen la misma distribución exponencial de tiempo de servicio, con parámetro μ. En este caso, sea n el número de servidores que *en este momento* prestan servicio y sea T_i el tiempo *que falta* para que el servidor i ($i = 1, 2, \ldots, n$) complete el servicio, que también tiene distribución exponencial con parámetro $\alpha_i = \mu$. Se puede concluir que U, el tiempo hasta la siguiente terminación de servicio para cualquier servidor, tiene una distribución exponencial con parámetro $\alpha = n\mu$. En efecto, el sistema de colas, *en este momento*, actúa como un sistema de un *solo* servidor, en el que los tiempos de servicio tienen una distribución exponencial con parámetro $n\mu$. En este capítulo se hará uso frecuente de esta implicación al analizar los modelos de varios servidores.

Cuando se usa esta propiedad, algunas veces es útil también determinar las probabilidades de *cuál* de las variables aleatorias exponenciales será la que tiene el valor mínimo. Por ejemplo, puede

quererse encontrar la probabilidad de que un servidor j en particular termine primero de servir a un cliente entre los n servidores ocupados. Es bastante sencillo (vea el problema 17.4-9) demostrar que esta probabilidad es proporcional al parámetro α_j. En particular, la probabilidad de que T_j resulte ser el menor de las n variables aleatorias es

$$P\{T_j = U\} = \frac{\alpha_j}{\sum_{i=1}^{n} \alpha_i}, \quad \text{para } j = 1, 2, \ldots, n.$$

Propiedad 4: Relación con la distribución de Poisson.

Suponga que el *tiempo* entre dos ocurrencias consecutivas de un tipo específico de evento (esto es, llegadas o terminación de servicio por un servidor siempre ocupado) tiene una distribución exponencial con parámetro α. La propiedad 4 está relacionada con la implicación resultante sobre la distribución de probabilidad del *número* de veces que ocurre este evento en un periodo dado. En particular, sea $X(t)$ el número de ocurrencias en el tiempo t ($t \geq 0$), donde el tiempo 0 es el instante en el que comienza la cuenta. La implicación es que

$$P\{X(t) = n\} = \frac{(\alpha t)^n e^{-\alpha t}}{n!}, \qquad \text{para } n = 0, 1, 2, \ldots;$$

es decir, $X(t)$ tiene una distribución de Poisson con parámetro αt. Por ejemplo, para $n = 0$,

$$P\{X(t) = 0\} = e^{-\alpha t},$$

que es exactamente la probabilidad que se obtuvo a partir de la distribución exponencial para que ocurra el *primer* evento después de un tiempo t. La media de la distribución de Poisson es

$$E\{X(t)\} = \alpha t,$$

de manera que el número esperado de eventos *por unidad de tiempo* es α. Por lo tanto, se dice que α es la *tasa media* a la que ocurren los eventos. Cuando se cuentan los eventos de manera continua, se dice que el proceso de conteo $\{X(t); t \geq 0\}$ es un **proceso de Poisson** con parámetro α (la tasa media).

Esta propiedad brinda información útil sobre la *terminación de servicio* cuando los tiempos de servicio tienen una distribución exponencial con parámetro μ. Esta información se obtiene al definir $X(t)$ como el número de servicios completos logrados por un servidor *siempre ocupado* en un tiempo transcurrido t, donde $\alpha = \mu$. En el caso de modelos de *múltiples* servidores, también se puede definir $X(t)$ como el número de terminaciones de servicio logradas por n servidores siempre ocupados en un tiempo transcurrido t, donde $\alpha = n\mu$.

Esta propiedad es útil en particular para describir el comportamiento probabilístico de las *llegadas* cuando los tiempos entre ellas siguen una distribución exponencial con parámetro λ. En este caso, $X(t)$ sería el *número* de llegadas en un tiempo transcurrido t, donde $\alpha = \lambda$ es la *tasa media de llegadas*. En consecuencia, las llegadas ocurren de acuerdo con un **proceso de entradas de Poisson** con parámetro λ. Este tipo de modelos se describe también con el supuesto de que tienen *llegadas de Poisson*.

Algunas veces se dice que las llegadas ocurren *aleatoriamente*, lo cual significa que suceden de acuerdo con un proceso de entradas de Poisson. Una interpretación intuitiva de este fenómeno es que cada periodo de longitud fija tiene la *misma* oportunidad de tener una llegada sin importar cuándo ocurrió la llegada anterior, como lo sugiere la siguiente propiedad.

Propiedad 5: Para todos los valores positivos de t, $P\{T \leq t + \Delta t \mid T > t\} \approx \alpha\, \Delta t$, para un Δt pequeño.

Todavía se interpreta T como el tiempo que pasa desde el último evento de cierto tipo (llegada o terminación de servicio) hasta el siguiente evento, y suponga que ha transcurrido un tiempo t sin que ocurra un evento. Se sabe, de acuerdo con la propiedad 2, que la probabilidad de que ocurra un evento dentro del siguiente intervalo, de longitud fija Δt, es una *constante* (que se identificará en el siguiente párrafo), sin que importe el tamaño de t. La propiedad 5 va más allá pues agrega

que, cuando el valor de Δt es pequeño, esta probabilidad constante se puede aproximar de manera muy cercana por $\alpha\ \Delta t$. Lo que es más, cuando se consideran distintos valores pequeños de Δt, esta probabilidad es, en esencia, *proporcional* a Δt, con factor de proporcionalidad igual α. En realidad, α es la *tasa media* a la cual ocurren los eventos (vea la propiedad 4), por lo que el *número esperado* de eventos en el intervalo de longitud Δt es *exactamente* $\alpha\ \Delta t$. La única razón por la que la probabilidad de que ocurra un evento difiere de este valor es la posibilidad de que ocurra *más de un evento*, lo cual tiene una probabilidad despreciable cuando Δt es pequeño.

Para comprobar de manera matemática por qué se cumple la propiedad 5, observe que el valor constante de la probabilidad (para un valor fijo de $\Delta t > 0$) es sólo

$$\begin{aligned} P\{T \le t + \Delta t \mid T > t\} &= P\{T \le \Delta t\} \\ &= 1 - e^{-\alpha\ \Delta t}, \end{aligned}$$

para cualquier $t \ge 0$. Por esta razón, como la expansión de la serie e^x para cualquier exponente x es

$$e^x = 1 + x + \sum_{n=2}^{\infty} \frac{x^n}{n!},$$

se concluye que

$$\begin{aligned} P\{T \le t + \Delta t \mid T > t\} &= 1 - 1 + \alpha\ \Delta t - \sum_{n=2}^{\infty} \frac{(-\alpha\ \Delta t)^n}{n!} \\ &\approx \alpha\ \Delta t, \qquad \text{para } \Delta t \text{ pequeño,}^3 \end{aligned}$$

porque los términos de la sumatoria se vuelven despreciables para valores de $\alpha\ \Delta t$ suficientemente pequeños.

Como en los modelos de colas T se pueden representar ya sea tiempos entre llegadas o tiempos de servicio, esta propiedad proporciona una aproximación conveniente de la probabilidad de que ocurra el evento de interés en el siguiente intervalo pequeño (Δt). También se puede hacer un análisis exacto basado en esta aproximación, tomando los límites apropiados cuando $\Delta t \to 0$.

Propiedad 6: No afecta agregar o desagregar.

Esta propiedad es importante para verificar que el *proceso de entrada* es de *Poisson*. Entonces, se describirá en estos términos, aunque también se aplica directamente a la distribución exponencial (tiempos entre llegadas exponenciales) debido a la propiedad 4.

Primero se considera agregar (combinar) varios procesos de entrada de Poisson en un proceso de entrada global. En particular, suponga que existen varios (n) tipos *diferentes* de clientes, en donde los clientes de cada tipo (tipo i) llegan de acuerdo a *un proceso de llegadas de Poisson* con parámetro λ_i ($i = 1, 2, \ldots, n$). Suponga que se trata de procesos de Poisson *independientes*. La propiedad dice que el proceso de entrada *agregado* (llegada de todos los clientes sin importar de qué tipo sean) también debe ser de Poisson, con parámetro (tasa de llegada) $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. En otras palabras, si se está en un proceso de Poisson *no afecta agregar*.

Esta parte de la propiedad se deduce de manera directa de las propiedades 3 y 4. Esta última implica que los tiempos entre llegadas de los clientes de tipo i tienen una distribución exponencial con parámetro λ_i. Para esta situación idéntica, ya se analizó en la propiedad 3 que esto implica que los tiempos entre llegadas de todos los clientes también deben tener una distribución exponencial, con parámetro $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. Si se usa la propiedad 4, de nuevo implica que el proceso de entrada agregado es de Poisson.

La segunda parte de la propiedad 6 ("no afecta desagregar") se refiere al caso contrario, en el que se sabe que el proceso de entrada *agregado* (que se obtuvo con la combinación de procesos de entrada para diferentes tipos de clientes) es de Poisson con parámetro λ. El cuestionamiento ahora

[3] De manera más precisa,

$$\lim_{\Delta t \to 0} \frac{P\{T \le t + \Delta t \mid T > t\}}{\Delta t} = \alpha.$$

concierne a la naturaleza del proceso de entrada *desagregado* de los tipos de clientes individuales. Si se supone que cada cliente que llega tiene una probabilidad *fija* p_i de pertenecer al tipo i ($i = 1, 2, \ldots, n$), con

$$\lambda_i = p_i\lambda \qquad \text{y} \qquad \sum_{i=1}^{n} p_i = 1,$$

la propiedad dice que el proceso de entrada de los clientes tipo i también debe ser de Poisson con parámetro λ_i. En otras palabras, si se está en un proceso de Poisson, *no afecta desagregar*.

Como ejemplo de la utilidad de esta segunda parte de la propiedad, considere la siguiente situación. Los clientes, sin hacer distinciones, llegan de acuerdo con un proceso de Poisson con parámetro λ. Cada cliente que llega tiene una probabilidad fija p de *desistir* (irse sin entrar al sistema de colas), de manera que la probabilidad de entrar es $(1 - p)$. Desde este punto de vista, existen dos tipos de clientes: aquellos que desisten y aquellos que entran al sistema. La propiedad dice que cada tipo llega de acuerdo con un proceso de Poisson, con parámetros $p\lambda$ y $(1 - p)\lambda$, respectivamente. Por lo tanto, al usar el último proceso de Poisson, los modelos de colas que suponen llegadas de Poisson todavía se pueden usar para analizar el funcionamiento del sistema de colas para aquellos clientes que entran al sistema.

Otro ejemplo en la sección Worked Examples del sitio en internet de este libro ilustra la aplicación de varias de las propiedades de la distribución exponencial que se presentan en esta sección.

17.5 PROCESO DE NACIMIENTO Y MUERTE

La mayor parte de los modelos elementales de colas suponen que las entradas (llegada de clientes) y las salidas (clientes que se van) del sistema ocurren de acuerdo con un *proceso de nacimiento y muerte*. Este importante proceso de teoría de probabilidad tiene aplicaciones en varias áreas. Sin embargo, en el contexto de la teoría de colas, el término **nacimiento** se refiere a la *llegada* de un nuevo cliente al sistema de colas, mientras que el término **muerte** se refiere a la *salida* del cliente servido. El *estado* del sistema en el tiempo t ($t \geq 0$), denotado por $N(t)$, es el número de clientes que hay en el sistema de colas en el tiempo t. El proceso de nacimiento y muerte describe en *términos probabilísticos* cómo cambia $N(t)$ al aumentar t. En general, sostiene que los nacimientos y muertes *individuales* ocurren de manera *aleatoria*, y que sus tasas medias de ocurrencia dependen del estado actual del sistema. De manera más precisa, los supuestos del proceso de nacimiento y muerte son los siguientes:

Supuesto 1. Dado $N(t) = n$, la distribución de probabilidad actual del tiempo que *falta* para el próximo *nacimiento* (llegada) es *exponencial* con parámetro λ_n ($n = 0, 1, 2, \ldots$).

Supuesto 2. Dado $N(t) = n$, la distribución de probabilidad actual del tiempo que *falta* para la próxima *muerte* (terminación de servicio) es *exponencial* con parámetro μ_n ($n = 1, 2, \ldots$).

Supuesto 3. La variable aleatoria del supuesto 1 (el tiempo que falta hasta el próximo *nacimiento*) y la variable aleatoria del supuesto 2 (el tiempo que falta hasta la siguiente *muerte*) son mutuamente independientes. La siguiente transición del estado del proceso es

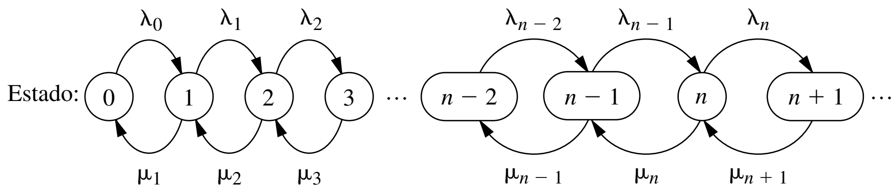
$$n \to n + 1 \text{(un solo nacimiento)}$$

o

$$n \to n - 1 \text{(una sola muerte)},$$

lo que depende de cuál de las dos variables es más pequeña.

En el caso de un sistema de colas, λ_n y μ_n representan, respectivamente, la *tasa media de llegada* y la *tasa media de terminaciones de servicio*, cuando hay n clientes en el sistema. En algunos sistemas de colas, los valores de las λ_n serán las mismas para todos los valores de n, y las μ_n también serán las mismas para toda n excepto para aquella n tan pequeña que el servidor esté desocupado (es decir, $n = 0$). Sin embargo, las λ_n y las μ_n también pueden variar en forma considerable con n para algunos sistemas de colas.



■ **FIGURA 17.4**
Diagrama de tasas del proceso de nacimiento y muerte.

Por ejemplo, una de las formas en las que λ_n puede ser diferente para valores distintos de n es si los clientes potenciales que llegan se pueden *perder* (rechazar la entrada al sistema) con mayor probabilidad a medida que n aumenta. De manera similar, μ_n puede ser diferente ante valores distintos de n debido a que existe una mayor probabilidad de que los clientes *renuncien* (se vayan sin haber sido servidos) a medida que aumenta el tamaño de la cola. Uno de los ejemplos de la sección Ejemplos resueltos del sitio en internet de este libro ilustra un sistema de líneas de espera donde existe tanto pérdida como renuncia. Entonces, este ejemplo demuestra cómo los resultados generales del proceso de nacimiento y muerte generan de manera directa varias medidas de desempeño de este sistema de colas.

Análisis del proceso de nacimiento y muerte

Como consecuencia de los supuestos 1 y 2, el proceso de nacimiento y muerte es un tipo especial de *cadena de Markov de tiempo continuo* (vea en la sección 16.8 la descripción de una cadena de Markov de tiempo continuo y sus propiedades y una introducción al procedimiento general para encontrar las probabilidades de estado estable que se aplicará en el resto de la sección). Los modelos de colas que se pueden representar mediante una cadena de Markov de tiempo continuo son mucho más manejables en el sentido analítico que cualquier otro modelo.

Como la propiedad 4 de la distribución exponencial implica que las λ_n y μ_n son tasas medias (vea la sección 17.4), estos supuestos se pueden resumir en el diagrama de tasas que se muestra en la figura 17.4. Las flechas de este diagrama muestran las únicas *transiciones* posibles en el estado del sistema (como lo especifica el supuesto 3) y el elemento junto a cada flecha es la tasa media de esa transición (según los supuestos 1 y 2) cuando el sistema se encuentra en el estado que hay en la base de la flecha.

Excepto en algunos casos especiales, el análisis del proceso de nacimiento y muerte es complicado cuando el sistema se encuentra en condición *transitoria*. Se han obtenido algunos resultados sobre esta distribución de probabilidad de $N(t)$ pero son demasiado complicados para darles un buen uso práctico. Por otro lado, es bastante directo derivar esta distribución *después* de que el sistema ha alcanzado la condición de *estado estable* (en caso de que pueda alcanzarla). Este desarrollo parte del diagrama de tasas, como se describe a continuación.

Considere cualquier estado particular n ($n = 0, 1, 2, \ldots$) del sistema. Suponga que en el tiempo 0 se inicia el conteo del número de veces que el sistema entra a este estado y el número de veces que sale de él, como se denota en seguida:

$E_n(t)$ = número de veces que el proceso entra al estado n hasta el tiempo t.

$L_n(t)$ = número de veces que el proceso sale del estado n hasta el tiempo t.

Como los dos tipos de eventos (entrar y salir) deben alternarse, estos dos números serán iguales o diferirán en sólo 1; es decir,

$$|E_n(t) - L_n(t)| \leq 1.$$

Al dividir ambos lados entre t y después hacer que $t \to \infty$ se obtiene

$$\left|\frac{E_n(t)}{t} - \frac{L_n(t)}{t}\right| \leq \frac{1}{t}, \qquad \text{entonces} \qquad \lim_{t\to\infty}\left|\frac{E_n(t)}{t} - \frac{L_n(t)}{t}\right| = 0.$$

Si se dividen $E_n(t)$ y $L_n(t)$ entre t se obtiene la *tasa real* (número de eventos por unidad de tiempo) a la que ocurren estos dos tipos de eventos, y cuando $t \to \infty$ se obtiene la *tasa media* (número esperado de eventos por unidad de tiempo):

$$\lim_{t\to\infty} \frac{E_n(t)}{t} = \text{tasa media a la que el proceso entra al estado } n.$$

$$\lim_{t\to\infty} \frac{L_n(t)}{t} = \text{tasa media a la que el proceso sale del estado } n.$$

Estos resultados conducen al siguiente principio clave:

Principio de tasa de entrada = tasa de salida. Para cualquier estado n ($n = 0, 1, 2, \ldots$) del sistema, la tasa media de entrada = tasa media de salida.

La ecuación que expresa este principio se llama **ecuación de balance** del estado n. Después de construir las ecuaciones de balance de todos los estados en términos de las probabilidades P_n *desconocidas*, se puede resolver este sistema de ecuaciones (más una ecuación que establezca que las probabilidades deben sumar 1) para encontrarlas.

A fin de ilustrar una ecuación de balance, considere el estado 0. El proceso entra a este estado *sólo* desde el estado 1. En consecuencia, la probabilidad de estado estable de encontrarse en el estado 1 (P_1) representa la proporción de tiempo que *es posible* que el proceso entre al estado 0. Dado que el proceso se encuentra en el estado 1, la tasa media de entrada al estado 0 es μ_1. (En otras palabras, para cada unidad acumulada de tiempo que el proceso pasa en el estado 1, el número esperado de veces que lo dejaría para entrar al estado 0 es μ_1.) Desde cualquier *otro* estado, esta tasa media es 0. Por lo tanto, la tasa media global a la que el proceso deja su estado actual para entrar al estado 0 (la *tasa media de entrada*) es

$$\mu_1 P_1 + 0(1 - P_1) = \mu_1 P_1.$$

Por el mismo razonamiento, la *tasa media de salida* debe ser $\lambda_0 P_0$, de manera que la ecuación de balance del estado 0 es

$$\mu_1 P_1 = \lambda_0 P_0.$$

En el caso de todos los demás estados, existen dos transiciones posibles, hacia adentro y hacia afuera del estado. Entonces, cada lado de las ecuaciones de balance de estos estados representa la *suma* de las tasas medias de las dos transiciones incluidas. Por lo demás, el razonamiento es igual que para el estado 0. Estas ecuaciones de balance se resumen en la tabla 17.1.

Observe que la primera ecuación de balance contiene dos variables (P_0 y P_1), las primeras dos ecuaciones contienen tres variables (P_0, P_1 y P_2), y así sucesivamente, de manera que siempre se tiene una variable "adicional". Por lo tanto, el procedimiento para resolver estas ecuaciones es despejar todas las variables en términos de una de ellas, entre las cuales la más conveniente es P_0. La primera ecuación se usa para despejar P_1 en términos de P_0; después se usa este resultado y la segunda ecuación para obtener P_2 en términos de P_0, etc. Al final, el requisito de que la suma de todas las probabilidades debe ser igual a 1 se puede usar para evaluar P_0.

TABLA 17.1 Ecuaciones de balance del proceso de nacimiento y muerte

Estado	Tasa de entrada = Tasa de salida
0	$\mu_1 P_1 = \lambda_0 P_0$
1	$\lambda_0 P_0 + \mu_2 P_2 = (\lambda_1 + \mu_1) P_1$
2	$\lambda_1 P_1 + \mu_3 P_3 = (\lambda_2 + \mu_2) P_2$
⋮	⋮
$n-1$	$\lambda_{n-2} P_{n-2} + \mu_n P_n = (\lambda_{n-1} + \mu_{n-1}) P_{n-1}$
n	$\lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1} = (\lambda_n + \mu_n) P_n$
⋮	⋮

Resultados del proceso de nacimiento y muerte

Al aplicar este procedimiento se obtienen los siguientes resultados:

Estado:

$$
\begin{array}{llllll}
0: & P_1 & = \dfrac{\lambda_0}{\mu_1}P_0 & & & \\
1: & P_2 & = \dfrac{\lambda_1}{\mu_2}P_1 + \dfrac{1}{\mu_2}(\mu_1 P_1 - \lambda_0 P_0) & = \dfrac{\lambda_1}{\mu_2}P_1 & = \dfrac{\lambda_1\lambda_0}{\mu_2\mu_1}P_0 \\
2: & P_3 & = \dfrac{\lambda_2}{\mu_3}P_2 + \dfrac{1}{\mu_3}(\mu_2 P_2 - \lambda_1 P_1) & = \dfrac{\lambda_2}{\mu_3}P_2 & = \dfrac{\lambda_2\lambda_1\lambda_0}{\mu_3\mu_2\mu_1}P_0 \\
\vdots & \vdots & & & \\
n-1: & P_n & = \dfrac{\lambda_{n-1}}{\mu_n}P_{n-1} + \dfrac{1}{\mu_n}(\mu_{n-1}P_{n-1} - \lambda_{n-2}P_{n-2}) & = \dfrac{\lambda_{n-1}}{\mu_n}P_{n-1} & = \dfrac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_1}P_0 \\
n: & P_{n+1} & = \dfrac{\lambda_n}{\mu_{n+1}}P_n + \dfrac{1}{\mu_{n+1}}(\mu_n P_n - \lambda_{n-1}P_{n-1}) & = \dfrac{\lambda_n}{\mu_{n+1}}P_n & = \dfrac{\lambda_n\lambda_{n-1}\cdots\lambda_0}{\mu_{n+1}\mu_n\cdots\mu_1}P_0 \\
\vdots & \vdots & & &
\end{array}
$$

Para simplificar la notación, sea

$$C_n = \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_1}, \qquad \text{para } n = 1, 2, \ldots,$$

y después se define $C_n = 1$ para $n = 0$. En este contexto, las probabilidades de estado estable son

$$P_n = C_n P_0, \qquad \text{para } n = 0, 1, 2, \ldots$$

El requisito

$$\sum_{n=0}^{\infty} P_n = 1$$

implica que

$$\left(\sum_{n=0}^{\infty} C_n\right) P_0 = 1,$$

así,

$$P_0 = \left(\sum_{n=0}^{\infty} C_n\right)^{-1}.$$

Cuando un modelo de líneas de espera se basa en el proceso de nacimiento y muerte, de manera que el estado del sistema n representa el número de clientes en el sistema de colas, las medidas clave de desempeño del sistema (L, L_q, W y W_q) se pueden obtener de inmediato después de calcular las P_n mediante las fórmulas anteriores. Las definiciones de L y L_q que se dieron en la sección 17.2 especifican que

$$L = \sum_{n=0}^{\infty} nP_n, \qquad L_q = \sum_{n=s}^{\infty} (n - s)P_n.$$

Lo que es más, las relaciones que se dieron en la sección 17.2 conducen a

$$W = \frac{L}{\bar{\lambda}}, \qquad W_q = \frac{L_q}{\bar{\lambda}},$$

donde $\overline{\lambda}$ es la tasa de llegadas *promedio* a largo plazo. Como λ_n es la tasa media de llegadas cuando el sistema se encuentra en el estado n ($n = 0, 1, 2, \ldots$) y P_n es la proporción de tiempo que el sistema está en este estado,

$$\overline{\lambda} = \sum_{n=0}^{\infty} \lambda_n P_n.$$

Varias de las expresiones que se acaban de presentar incluyen sumas con un número infinito de términos. Por fortuna, estas sumas tienen soluciones analíticas de muchos casos especiales interesantes[4] como se verá en la siguiente sección. En otros casos, se puede aproximar al sumar un número finito de términos en una computadora.

Estos resultados de estado estable se desarrollaron bajo el supuesto de que los parámetros λ_n y μ_n tienen valores tales que el proceso, en realidad, puede *alcanzar* la condición de estado estable. Este supuesto *siempre* se cumple si $\lambda_n = 0$ para algún valor de n mayor que el estado inicial, de forma que sólo son posibles un número finito de estados (aquellos menores que esta n). También se cumple *siempre* cuando λ y μ están definidas (vea "Terminología y notación" en la sección 17.2) y $\rho = \lambda/(s\mu) < 1$. *No* se cumple si $\sum_{n=1}^{\infty} C_n = \infty$.

En la sección 17.6 se describen varios modelos de colas que son casos especiales del proceso de nacimiento y muerte. Los resultados de estado estable generales dados en los recuadros se utilizarán una y otra vez para obtener los resultados específicos para estos modelos.

17.6 MODELOS DE COLAS BASADOS EN EL PROCESO DE NACIMIENTO Y MUERTE

Como se puede asignar cualquier valor no negativo a cada una de las tasas medias λ_0, λ_1, ... y μ_l, μ_2,... del proceso de nacimiento y muerte, se cuenta con una gran flexibilidad para modelar un sistema de colas. Los modelos que acaso sean los que más se usan en teoría de colas se basan directamente en este proceso. De acuerdo con los supuestos 1 y 2 (y la propiedad 4 de la distribución exponencial), se dice que estos modelos tienen **entradas de Poisson y tiempos de servicio exponencial**. Los modelos difieren sólo en los supuestos sobre cómo cambian las λ_n y las μ_n según el estado n. En esta sección se presentarán tres modelos de tres tipos importantes de sistemas de colas.

Modelo *M/M/s*

Como se describió en la sección 17.2, el modelo *M/M/s* supone que todos los *tiempos entre llegadas* son independientes e idénticamente distribuidos de acuerdo con una distribución exponencial (es decir, el proceso de entrada es de Poisson), que todos los *tiempos de servicio* son independientes e idénticamente distribuidos de acuerdo con otra distribución exponencial y que el número de servidores es s (cualquier entero positivo). En consecuencia, este modelo es sólo un caso especial del proceso de nacimiento y muerte cuando la *tasa media de llegadas* al sistema de colas y la *tasa media de servicio por servidor ocupado* son constantes (λ y μ, respectivamente) e independientes del estado del sistema. Cuando el sistema tiene sólo *un servidor* ($s = 1$), la implicación es que los parámetros del proceso de nacimiento y muerte son $\lambda_n = \lambda (n = 0, 1, 2, ...)$ y $\mu_n = \mu (n = 1, 2, ...)$. En la figura 17.5*a* se muestra el diagrama de tasas resultante.

Sin embargo, cuando el sistema tiene *varios servidores* ($s > 1$), no es tan sencillo expresar μ_n, como se explica en seguida.

Tasa de servicio de sistema: La tasa del servicio del sistema μ_n representa la tasa media de los servicios terminados de todo el sistema de colas cuando existen n clientes en él. En

[4] Estas soluciones se basan en los siguientes resultados de la suma de cualquier serie geométrica:

$$\sum_{n=0}^{N} x^n = \frac{1 - x^{N+1}}{1 - x}, \qquad \text{para cualquier } x \neq 1,$$

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1 - x}, \qquad \text{si } |x| < 1.$$

el caso de múltiples servidores y $n > 1$, μ_n no es lo mismo que μ, la tasa media de servicio por servidor ocupado. En lugar de eso,

$$\mu_n = n\mu \quad \text{cuando } n \leq s,$$
$$\mu_n = s\mu \quad \text{cuando } n \geq s.$$

Mediante el uso de estas fórmulas de μ_n, el diagrama con las tasas del proceso nacimiento-muerte que se muestra en la figura 17.4 se reduce a los diagramas de las tasas que se muestran en la figura 17.5 del modelo *M/M/s*.

Cuando $s\mu$ excede la tasa media de llegadas λ, es decir, cuando

$$\rho = \frac{\lambda}{s\mu} < 1,$$

un sistema de colas que se ajuste a este modelo tarde o temprano alcanzará la condición de estado estable. En esta situación se pueden aplicar directamente los resultados de estado estable que se obtuvieron en la sección 17.5 al proceso de nacimiento y muerte. Estos resultados se simplifican mucho para este modelo y proporcionan expresiones cortas de P_n, L, L_q, etc., como se mostrará a continuación.

Resultados en el caso de un servidor (*M/M/*1). Para $s = 1$, los factores C_n del proceso de nacimiento y muerte se reducen a

$$C_n = \left(\frac{\lambda}{\mu}\right)^n = \rho^n, \qquad \text{para } n = 0, 1, 2, \ldots$$

Por lo tanto,

$$P_n = \rho^n P_0, \qquad \text{para } n = 0, 1, 2, \ldots,$$

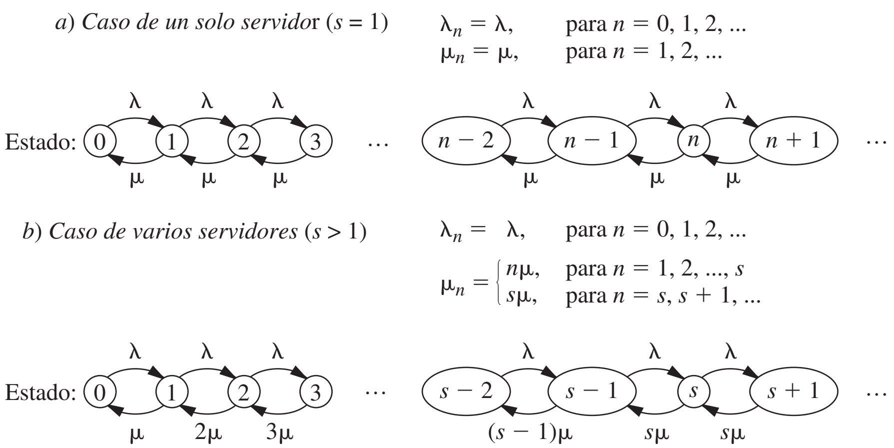
donde

$$P_0 = \left(\sum_{n=0}^{\infty} \rho^n\right)^{-1}$$
$$= \left(\frac{1}{1-\rho}\right)^{-1}$$
$$= 1 - \rho.$$

Así

$$P_n = (1 - \rho)\rho^n, \qquad \text{para } n = 0, 1, 2, \ldots.$$

■ **FIGURA 17.5**
Diagrama de tasas del modelo *M/M/s*.



Recuadro de aplicación

KeyCorp, una compañía cuya casa matriz está en Cleveland, Ohio, se encuentra dentro de las 500 compañías de Fortune. Es la decimotercera compañía bancaria más grande de Estados Unidos. Emplea a 19 000 personas, tiene activos por 93 000 millones de dólares y ganancias anuales por 6 700 millones de dólares. La compañía se enfoca en prestar servicios bancarios a los consumidores y cuenta con 2.4 millones de clientes repartidos en 1 300 sucursales y oficinas afiliadas.

Para impulsar el crecimiento de sus negocios, la gerencia de KeyCorp puso en marcha un amplio estudio de IO para determinar la forma de mejorar el servicio a los clientes (el cual se define como la reducción del tiempo de espera para recibir el servicio), y a la vez conservar los costos de personal en un rango moderado. Se fijó una objetivo en cuanto a la calidad del servicio que consistía en que el tiempo de espera de los clientes debería ser menor a 5 minutos.

La herramienta clave para analizar este problema fue el *modelo de colas M/M/s*, el cual demostró ser el más conveniente para esta aplicación. Para aplicarlo, se recabaron datos que revelaron que el tiempo de servicio promedio que se requería para procesar un cliente era muy elevado, del orden de 246 segundos. Con este tiempo promedio de servicio y tasa de arribo en la media típicamente, el modelo indicó que era necesario un incremento de 30% del número de personas que brindan atención a los clientes con el fin de cumplir con el objetivo de calidad del servicio. Esta opción extremadamente costosa llevó a la alta gerencia a concluir que era necesario llevar a cabo una amplia campaña para reducir de manera drástica el tiempo de servicio promedio mediante la reingeniería de las sesiones con los clientes y del fomento de una mejor administración del personal. En un periodo de tres años, esta campaña dio como resultado una reducción del tiempo de servicio promedio hasta llegar a 115 segundos. La aplicación frecuente del modelo *M/M/s* reveló la forma en que la meta de calidad del servicio puede sobrepasarse de manera significativa, a la vez que se pueden reducir los niveles de personal mediante una programación optimizada de los empleados en las diferentes sucursales del banco.

El resultado neto se ha reflejado en *ahorros de casi* **20 millones de dólares** *por año con un servicio significativamente mejorado* que permite que 96% de los clientes tengan que esperar menos de 5 minutos. Esta mejora se extendió a toda la compañía ya que el porcentaje de sucursales que cumple el objetivo de calidad de servicio aumentó de 42 a 94 por ciento. Diferentes estudios también confirman un incremento significativo en la satisfacción del cliente.

Fuente: S. K. Kotha, M. P. Barnum y D. A. Bowen: "KeyCorp Service Excellence Management System", en *Interfaces*, **26**(1): 54-74, enero-febrero de 1996. (En el sitio en internet de este libro www.mhhe.com/hillier se proporciona una liga hacia este artículo.)

En consecuencia,

$$
\begin{aligned}
L &= \sum_{n=0}^{\infty} n(1-\rho)\rho^n \\
&= (1-\rho)\rho \sum_{n=0}^{\infty} \frac{d}{d\rho}(\rho^n) \\
&= (1-\rho)\rho \frac{d}{d\rho}\left(\sum_{n=0}^{\infty} \rho^n\right) \\
&= (1-\rho)\rho \frac{d}{d\rho}\left(\frac{1}{1-\rho}\right) \\
&= \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}.
\end{aligned}
$$

En forma similar,

$$
\begin{aligned}
L_q &= \sum_{n=1}^{\infty} (n-1)P_n \\
&= L - 1(1-P_0) \\
&= \frac{\lambda^2}{\mu(\mu-\lambda)}.
\end{aligned}
$$

Cuando $\lambda \geq \mu$, esto es, la tasa media de llegadas excede la tasa media de servicio, la solución anterior "no sirve" (puesto que la suma para calcular P_0 diverge). En este caso, la cola "explota" y crece sin límite. Si el sistema de colas comienza a operar sin clientes presentes, puede ser que el

servidor se dé abasto con los que llegan durante un periodo corto, pero a la larga le será imposible. (Aun cuando $\lambda = \mu$, el número *esperado* de clientes en el sistema crecerá sin límite y con lentitud a través del tiempo y, aunque siempre es posible un regreso temporal a no tener clientes, las probabilidades de tener números grandes de clientes crecen en forma significativa con el tiempo.)

Si se supone de nuevo que $\lambda < \mu$, se puede obtener la distribución de probabilidad del *tiempo de espera en el sistema* (se *incluye* el tiempo de servicio) $\mathcal{W}$ de una llegada aleatoria cuando la disciplina de la cola es primero en entrar, primero en salir. Si esta llegada encuentra n clientes en el sistema, tendrá que esperar $n + 1$ tiempos de servicio exponenciales, inclusive el propio. (Para el cliente que en la actualidad se encuentra en servicio, recuerde la propiedad de falta de memoria de la distribución exponencial que se presentó en la sección 17.4.) Por lo tanto, sean $T_1, T_2, \ldots$ las variables aleatorias independientes de los tiempos de servicio que tienen una distribución exponencial con parámetro μ, y sea

$$S_{n+1} = T_1 + T_2 + \cdots + T_{n+1}, \qquad \text{para } n = 0, 1, 2, \ldots,$$

de manera que S_{n+1} representa el tiempo de espera *condicional*, dado que hay n clientes en el sistema. Como se analiza en la sección 17.7, se sabe que S_{n+1} tiene *distribución de Erlang*.[5] Como la probabilidad de que una llegada aleatoria encuentre n clientes en el sistema es P_n, se concluye que

$$P\{\mathcal{W} > t\} = \sum_{n=0}^{\infty} P_n P\{S_{n+1} > t\},$$

lo que después de una manipulación algebraica considerable (vea el problema 17.6-17) se reduce a

$$P\{\mathcal{W} > t\} = e^{-\mu(1-\rho)t}, \qquad \text{para } t \geq 0.$$

La conclusión sorprendente es que $\mathcal{W}$ tiene una distribución *exponencial* con parámetro igual a $\mu(1 - \rho)$. Por lo tanto,

$$W = E(\mathcal{W}) = \frac{1}{\mu(1-\rho)} = \frac{1}{\mu - \lambda}.$$

Estos resultados *incluyen* el tiempo de servicio en el tiempo de espera. En algunos contextos (por ejemplo, el problema de la sala de emergencia del Hospital General descrito en la sección 17.1), el tiempo de espera más importante es hasta que comienza el servicio. Considere el *tiempo de espera en la cola* (*excluya* el tiempo de servicio) $\mathcal{W}_q$ de la llegada aleatoria cuando la disciplina de la cola es primero en llegar, primero en salir. Si esta llegada no encuentra clientes en el sistema, se le sirve de inmediato, de manera que

$$P\{\mathcal{W}_q = 0\} = P_0 = 1 - \rho.$$

Si encuentra $n > 0$ clientes, entonces tendrá que esperar n tiempos de servicio exponenciales hasta que su propio servicio comience, de forma que

$$\begin{aligned} P\{\mathcal{W}_q > t\} &= \sum_{n=1}^{\infty} P_n P\{S_n > t\} \\ &= \sum_{n=1}^{\infty} (1-\rho)\rho^n P\{S_n > t\} \\ &= \rho \sum_{n=0}^{\infty} P_n P\{S_{n+1} > t\} \\ &= \rho P\{\mathcal{W} > t\} \\ &= \rho e^{-\mu(1-\rho)t}, \qquad \text{para } t \geq 0. \end{aligned}$$

[5] Fuera del contexto de teoría de colas, esta distribución se conoce como *distribución gamma*.

Observe que W_q no tiene en realidad una distribución exponencial porque $P\{\mathcal{W}_q = 0\} > 0$. Sin embargo, la distribución *condicional* de $\mathcal{W}_q$, dado que $\mathcal{W}_q > 0$, no tiene una distribución exponencial con parámetro $\mu(1-\rho)$, como lo tiene $\mathcal{W}$, porque

$$P\{\mathcal{W}_q > t \mid \mathcal{W}_q > 0\} = \frac{P\{\mathcal{W}_q > t\}}{P\{\mathcal{W}_q > 0\}} = e^{-\mu(1-\rho)t}, \qquad \text{para } t \geq 0.$$

Al obtener la media de esta distribución (no condicional) de W_q (o aplicar una de las dos fórmulas $L_q = \lambda W_q$ o $W_q = W - 1/\mu$), resulta

$$W_q = E(\mathcal{W}_q) = \frac{\lambda}{\mu(\mu - \lambda)}.$$

Si desea ver **otro ejemplo** que aplica el modelo *M/M/*1, en la sección Worked Examples del sitio en internet de este libro se presenta uno de ellos para determinar qué tipo de equipo de manejo de materiales debe comprar una compañía.

Resultados del caso de varios servidores ($s > 1$). Cuando $s > 1$, los factores C_n se convierten en

$$C_n = \begin{cases} \dfrac{(\lambda/\mu)^n}{n!} & \text{para } n = 1, 2, \ldots, s \\[2ex] \dfrac{(\lambda/\mu)^s}{s!}\left(\dfrac{\lambda}{s\mu}\right)^{n-s} = \dfrac{(\lambda/\mu)^n}{s!\,s^{n-s}} & \text{para } n = s, s+1, \ldots. \end{cases}$$

En consecuencia, si $\lambda < s\mu$ [de manera que $\rho = \lambda/(s\mu) < 1$], entonces

$$\begin{aligned} P_0 &= 1 \Bigg/ \left[1 + \sum_{n=1}^{s-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!} \sum_{n=s}^{\infty} \left(\frac{\lambda}{s\mu}\right)^{n-s}\right] \\ &= 1 \Bigg/ \left[\sum_{n=0}^{s-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!} \frac{1}{1 - \lambda/(s\mu)}\right], \end{aligned}$$

donde el término para $n = 0$ en la última suma lleva al valor correcto de 1 debido a la convención de que $n! = 1$ cuando $n = 0$. Estos factores C_n dan también

$$P_n = \begin{cases} \dfrac{(\lambda/\mu)^n}{n!} P_0 & \text{si } 0 \leq n \leq s \\[2ex] \dfrac{(\lambda/\mu)^n}{s!\,s^{n-s}} P_0 & \text{si } n \geq s. \end{cases}$$

Más aún,

$$\begin{aligned} L_q &= \sum_{n=s}^{\infty} (n - s) P_n \\ &= \sum_{j=0}^{\infty} j P_{s+j} \\ &= \sum_{j=0}^{\infty} j \frac{(\lambda/\mu)^s}{s!} \rho^j P_0 \\ &= P_0 \frac{(\lambda/\mu)^s}{s!} \rho \sum_{j=0}^{\infty} \frac{d}{d\rho} (\rho^j) \\ &= P_0 \frac{(\lambda/\mu)^s}{s!} \rho \frac{d}{d\rho} \left(\sum_{j=0}^{\infty} \rho^j\right) \end{aligned}$$

$$= P_0\frac{(\lambda/\mu)^s}{s!}\rho\,\frac{d}{d\rho}\left(\frac{1}{1-\rho}\right)$$

$$= \frac{P_0(\lambda/\mu)^s\rho}{s!(1-\rho)^2};$$

$$W_q = \frac{L_q}{\lambda};$$

$$W = W_q + \frac{1}{\mu};$$

$$L = \lambda\left(W_q + \frac{1}{\mu}\right) = L_q + \frac{\lambda}{\mu}.$$

En la figura 17.6 se muestra cómo cambian L con ρ para diferentes valores de s.

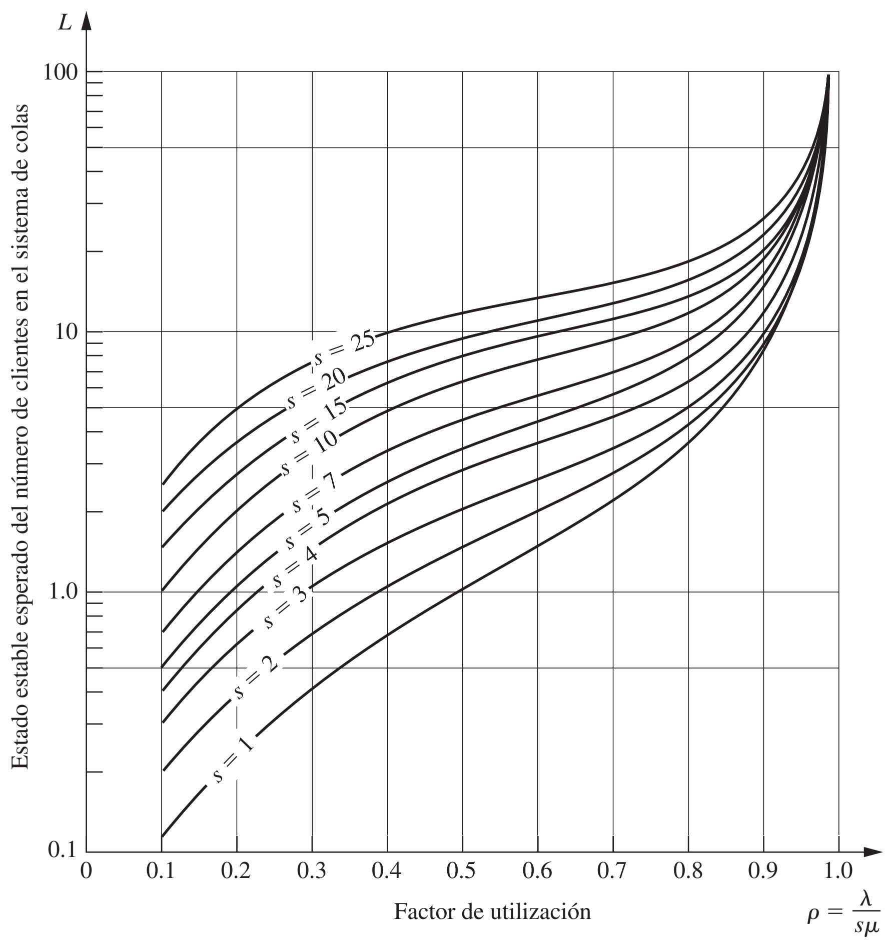
El método de un solo servidor para encontrar la distribución de probabilidad de los tiempos de espera se puede extender al caso de varios servidores. Al aplicarlo se obtiene[6] (para $t \geq 0$)

$$P\{\mathcal{W} > t\} = e^{-\mu t}\left[\frac{1 + P_0(\lambda/\mu)^s}{s!(1-\rho)}\left(\frac{1 - e^{-\mu t(s-1-\lambda/\mu)}}{s-1-\lambda/\mu}\right)\right]$$

y

$$P\{\mathcal{W}_q > t\} = (1 - P\{\mathcal{W}_q = 0\})e^{-s\mu(1-\rho)t},$$

■ **FIGURA 17.6**
Valores de L del modelo $M/M/s$ (sección 17.6).



[6] Cuando $s - 1 - \lambda/\mu = 0$, $(1 - e^{-\mu t(s-1-\lambda/\mu)})/(s - 1 - \lambda/\mu)$ debe sustituirse por μt.

donde

$$P\{\mathcal{W}_q = 0\} = \sum_{n=0}^{s-1} P_n.$$

Las fórmulas anteriores de las distintas medidas de desempeño (incluyendo las P_n) son bastante tediosas para hacer los cálculos a mano. Sin embargo, el archivo de Excel de este capítulo en OR Courseware incluye una plantilla Excel que realiza todos los cálculos simultáneos para cualesquiera valores de t, s, λ y μ dado que $\lambda < s\mu$.

Si $\lambda \geq s\mu$, de forma que si la tasa media de llegadas excede a la tasa media máxima de servicio, la cola crece sin límite y las soluciones de estado estable anteriores no se pueden aplicar.

Ejemplo del Hospital General con el modelo *M/M/s*. En el problema de la sala de emergencia del Hospital General (vea la sección 17.1), el ingeniero administrador ha concluido que los casos de emergencia llegan casi de manera aleatoria (*proceso de entrada de Poisson*), por lo que los tiempos entre llegadas tienen una distribución exponencial. También llegó a la conclusión de que el tiempo que necesita un médico para atender a los pacientes sigue aproximadamente una *distribución exponencial*. Con base en este contexto, eligió el modelo *M/M/s* para hacer un estudio preliminar de este sistema de colas.

Al proyectar los datos disponibles del turno de la tarde al año próximo, estima que los pacientes llegarán a una tasa *promedio* de uno cada media hora. Un médico requiere un promedio de 20 minutos para atender al paciente. Si se usa una hora como unidad de tiempo,

$$\frac{1}{\lambda} = \frac{1}{2} \text{ horas por cliente}$$

y

$$\frac{1}{\mu} = \frac{1}{3} \text{ horas por cliente,}$$

TABLA 17.2 Resultados de estado estable del modelo *M/M/s* del problema del Hospital General

	$s = 1$	$s = 2$
ρ	$\frac{2}{3}$	$\frac{1}{3}$
P_0	$\frac{1}{3}$	$\frac{1}{2}$
P_1	$\frac{2}{9}$	$\frac{1}{3}$
P_n para $n \geq 2$	$\frac{1}{3}\left(\frac{2}{3}\right)^n$	$\left(\frac{1}{3}\right)^n$
L_q	$\frac{4}{3}$	$\frac{1}{12}$
L	2	$\frac{3}{4}$
W_q	$\frac{2}{3}$ horas	$\frac{1}{24}$ horas
W	1 hora	$\frac{3}{8}$ horas
$P\{\mathcal{W}_q > 0\}$	0.667	0.167
$P\left\{\mathcal{W}_q > \frac{1}{2}\right\}$	0.404	0.022
$P\{\mathcal{W}_q > 1\}$	0.245	0.003
$P\{\mathcal{W}_q > t\}$	$\frac{2}{3}e^{-t}$	$\frac{1}{6}e^{-4t}$
$P\{\mathcal{W} > t\}$	e^{-t}	$\frac{1}{2}e^{-3t}(3 - e^{-t})$

de manera que

$$\lambda = 2 \text{ clientes por hora}$$

y

$$\mu = 3 \text{ clientes por hora.}$$

Las dos alternativas bajo consideración son: continuar con un solo médico durante este turno ($s = 1$) o agregar un segundo médico ($s = 2$). En ambos casos,

$$\rho = \frac{\lambda}{s\mu} < 1,$$

de forma que el sistema debe acercarse a la condición de estado estable. (En realidad, como λ varía un poco durante los otros turnos, el sistema nunca alcanzará verdaderamente la condición de estado estable, pero el ingeniero administrador piensa que los resultados correspondientes proporcionarán una buena aproximación.) Por lo tanto, usa las ecuaciones anteriores para obtener los resultados que se muestran en la tabla 17.2.

Con base en estos resultados, concluye en forma tentativa que para el siguiente año sería inadecuado un solo médico para brindar atención con relativa prontitud, lo que es necesario en la sala de emergencias de un hospital. Más adelante se verá (sección 17.8) cómo el ingeniero rectifica su conclusión cuando aplica otro modelo de colas que le proporciona una mejor representación de algunos aspectos cruciales del sistema de líneas de espera real.

Se puede ver **otro ejemplo** de una aplicación del modelo *M/M/*1 en la sección Worked Examples del sitio en internet de este libro, donde el problema consiste en determinar si tres empleados de un restaurante de comida rápida deben trabajar juntos como un servidor rápido o por separado como tres servidores considerablemente más lentos.

Variación de cola finita al modelo *M/M/s* (llamado modelo *M/M/s/K*)

En el análisis de la sección 17.2 se mencionó que los sistemas de colas a veces tienen una *cola finita*; esto es, no se permite que el número de clientes en el sistema exceda un número especificado (denotado por K), por lo que la capacidad de la cola es $K - s$. A cualquier cliente que llega cuando la cola está "llena" se le niega la entrada al sistema y este cliente lo deja para siempre. Desde el punto de vista del proceso de nacimiento y muerte, la tasa media de entrada al sistema se hace cero en estos momentos. Por lo mismo, la única modificación necesaria en el modelo *M/M/s* para introducir una cola finita es cambiar los parámetros λ_n a

$$\lambda_n = \begin{cases} \lambda & \text{para } n = 0, 1, 2, \ldots, K-1 \\ 0 & \text{para } n \geq K. \end{cases}$$

Como $\lambda_n = 0$ para algunos valores de n, un sistema de colas que se ajuste a este modelo alcanzará en algún momento la condición de estado estable, aun cuando $\rho = \lambda/s\mu \geq 1$.

Por lo general este modelo se etiqueta como *M/M/s/K*, donde la presencia del cuarto símbolo lo distingue del modelo *M/M/s*. La única diferencia en la formulación de los dos modelos es que K es finito en el modelo *M/M/s/K* y $K = \infty$ en el modelo *M/M/s*.

La interpretación física usual del modelo *M/M/s/K* es que se cuenta con un *espacio limitado de espera* que admite un máximo de K clientes en el sistema. Por ejemplo, en el problema de la sala de emergencias del Hospital General, este sistema en realidad tendría una cola finita si sólo hubiera K camillas para los pacientes y si la política fuera mandar a otro hospital a aquellos que llegan cuando no hay lugares disponibles.

Otra interpretación posible es que los clientes que llegan dejarán el sistema y se "irán a otra parte" cuando encuentren demasiados clientes (K) en el sistema antes que ellos, pues no están dispuestos a soportar una larga espera. Este fenómeno de desistir es bastante común en los sistemas de

servicio comercial, pero existen otros modelos disponibles (por ejemplo, vea el problema 17.5-5) que se ajustan mejor a esta situación.

El diagrama de tasas de este modelo es idéntico al que se muestra en la figura 17.5 para el modelo *M/M/s*, excepto que se detiene en el estado *K*.

Resultados en el caso de un servidor (*M/M/1/K*). En este caso,

$$C_n = \begin{cases} \left(\dfrac{\lambda}{\mu}\right)^n = \rho^n & \text{para } n = 0, 1, 2, \ldots, K \\ 0 & \text{para } n > K. \end{cases}$$

Entonces, para $\rho \neq 1$.[7]

$$\begin{aligned} P_0 &= \frac{1}{\sum_{n=0}^{K} (\lambda/\mu)^n} \\ &= 1 \Big/ \left[\frac{1-(\lambda/\mu)^{K+1}}{1-\lambda/\mu}\right] \\ &= \frac{1-\rho}{1-\rho^{K+1}}, \end{aligned}$$

de manera que

$$P_n = \frac{1-\rho}{1-\rho^{K+1}}\rho^n, \qquad \text{para } n = 0, 1, 2, \ldots, K.$$

Entonces,

$$\begin{aligned} L &= \sum_{n=0}^{K} nP_n \\ &= \frac{1-\rho}{1-\rho^{K+1}}\rho\sum_{n=0}^{K}\frac{d}{d\rho}(\rho^n) \\ &= \frac{1-\rho}{1-\rho^{K+1}}\rho\frac{d}{d\rho}\left(\sum_{n=0}^{K}\rho^n\right) \\ &= \frac{1-\rho}{1-\rho^{K+1}}\rho\frac{d}{d\rho}\left(\frac{1-\rho^{K+1}}{1-\rho}\right) \\ &= \rho\frac{-(K+1)\rho^K + K\rho^{K+1} + 1}{(1-\rho^{K+1})(1-\rho)} \\ &= \frac{\rho}{1-\rho} - \frac{(K+1)\rho^{K+1}}{1-\rho^{K+1}}. \end{aligned}$$

Como es usual (cuando $s = 1$),

$$L_q = L - (1 - P_0).$$

Note que los resultados anteriores no exigen que $\lambda < \mu$ (esto es, que $\rho < 1$.

Cuando $\rho < 1$, se puede verificar que el segundo término de la última expresión de L converge hacia 0 cuando $K \to \infty$, por lo que, sin duda, todos los resultados anteriores convergen hacia los resultados correspondientes que se obtuvieron antes en el caso del modelo *M/M/*1.

La distribución de los tiempos de espera se puede deducir si se utiliza el mismo razonamiento que para el modelo *M/M/*1 (vea el problema 17.6-28). Sin embargo, no se han obtenido expresiones sencillas en este caso, por lo que es preciso efectuar los cálculos en una computadora. Por fortuna, aun cuando en este modelo $L \neq \lambda W$ y $L_q \neq \lambda W_q$ puesto que las λ_n no son iguales para toda n (vea el

[7] Si $\rho = 1$, entonces $P_n = 1/(K + 1)$ para $n = 0, 1, 2, \ldots, K$, de forma que $L = K/2$.

final de la sección 17.2), sí se pueden obtener los tiempos de espera *esperados* de los clientes que llegan al sistema, en forma directa de las expresiones que se dieron al final de la sección 17.5:

$$W = \frac{L}{\bar{\lambda}}, \qquad W_q = \frac{L_q}{\bar{\lambda}},$$

donde

$$\begin{aligned}\bar{\lambda} &= \sum_{n=0}^{\infty} \lambda_n P_n \\ &= \sum_{n=0}^{K-1} \lambda P_n \\ &= \lambda(1 - P_K).\end{aligned}$$

Resultados en el caso de varios servidores ($s > 1$). Debido a que este modelo no permite más de K clientes en el sistema, K es el número máximo de servidores que pueden tenerse. Suponga que $s \leq K$. En este caso, C_n se expresa como

$$C_n = \begin{cases} \dfrac{(\lambda/\mu)^n}{n!} & \text{para } n = 0, 1, 2, \ldots, s \\ \dfrac{(\lambda/\mu)^s}{s!}\left(\dfrac{\lambda}{s\mu}\right)^{n-s} = \dfrac{(\lambda/\mu)^n}{s!\,s^{n-s}} & \text{para } n = s, s+1, \ldots, K \\ 0 & \text{para } n > K. \end{cases}$$

Así,

$$P_n = \begin{cases} \dfrac{(\lambda/\mu)^n}{n!} P_0 & \text{para } n = 1, 2, \ldots, s \\ \dfrac{(\lambda/\mu)^n}{s!\,s^{n-s}} P_0 & \text{para } n = s, s+1, \ldots, K \\ 0 & \text{para } n > K, \end{cases}$$

donde

$$P_0 = 1 \Bigg/ \left[\sum_{n=0}^{s} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!} \sum_{n=s+1}^{K} \left(\frac{\lambda}{s\mu}\right)^{n-s}\right].$$

(Estas fórmulas aplican la convención de que $n! = 1$ cuando $n = 0$.)
Si se adapta a este caso la derivación de L_q del modelo *M/M/s*, se llega a

$$L_q = \frac{P_0(\lambda/\mu)^s \rho}{s!(1-\rho)^2}\,[1 - \rho^{K-s} - (K-s)\rho^{K-s}(1-\rho)],$$

donde $\rho = \lambda/(s\mu)$.[8] Se puede demostrar que

$$L = \sum_{n=0}^{s-1} nP_n + L_q + s\left(1 - \sum_{n=0}^{s-1} P_n\right).$$

Y W y W_q se obtienen a partir de estas cantidades, como se mostró en el caso de un solo servidor.

El archivo de Excel de este capítulo incluye una plantilla para calcular las medidas de desempeño anteriores (inclusive las P_n) de este modelo.

Un caso especial e interesante de este modelo es cuando $K = s$, de manera que la capacidad de la cola es $K - s = 0$. En este caso, los clientes que llegan cuando todos los servidores están ocupados dejan el sistema de inmediato y se pierden. Esto ocurrirá, por ejemplo, en una red telefónica con s líneas troncales de manera que cuando todas estas líneas están ocupadas, quien llama obtiene una

---
[8] Si $\rho = 1$, es necesario aplicar la regla de L'Hôpital dos veces a esta expresión de L_q. De otra manera, todos estos resultados se cumplen para toda $\rho > 0$. La razón para que este sistema de colas alcance la condición de estado estable aun cuando $\rho \geq 1$ es que $\lambda_n = 0$ para toda $n \geq K$, de modo que el número de clientes en el sistema no puede seguir creciendo en forma indefinida.

señal de ocupado y cuelga. Este tipo de sistema (un "sistema de colas" sin cola) se conoce como *sistema de pérdidas Erlang*, porque el primero que lo estudió a principios del siglo XX fue A. K. Erlang, un ingeniero en telefonía holandés, a quien se considera el fundador de la teoría de colas.

Ahora resulta común que los sistemas telefónicos de un centro de llamadas proporcionen algunas líneas extras que colocan a la persona que llama en espera, pero los clientes adicionales obtendrán una señal de ocupado. Dicho sistema también se ajusta a este modelo, donde $(K - s)$ es el número de líneas adicionales que colocan las llamadas en espera. Uno de los ejemplos de la sección Worked Examples del sitio en internet de este libro ilustra la aplicación de este modelo a un sistema como el descrito.

Variación de fuente de entrada finita al modelo *M/M/s*

Ahora suponga que la única diferencia con el modelo *M/M/s* (según se definió en la sección 17.2) es que la *fuente de entrada* está *limitada*; es decir, el tamaño de la *población potencial es finito*. En este caso, sea N el tamaño de esa población. Cuando el número de clientes en el sistema de colas es n ($n = 0, 1, 2, \ldots, N$), existen sólo $N - n$ clientes *potenciales* restantes en la fuente de entrada.

La aplicación más importante de este modelo es el problema de reparación de máquinas, en el que se asigna a uno o más técnicos la responsabilidad de mantener en operación cierto grupo de N máquinas dando servicio a cada una de las que se descomponen. (El ejemplo que se presentó al final de la sección 16.8 ilustra esta aplicación cuando se usa el procedimiento general para resolver cualquier *cadena de Markov de tiempo continuo* en lugar de las fórmulas específicas disponibles para el proceso de nacimiento y muerte.) Se considera que un técnico de mantenimiento es un servidor individual en el sistema de colas si trabaja en forma independiente en máquinas diferentes, mientras que los miembros de una cuadrilla completa se toman como un servidor si trabajan unidos en cada máquina. Las máquinas constituyen la población potencial. Cada una se considera un cliente en el sistema de colas cuando está descompuesta en espera de ser reparada, mientras que cuando está en operación está fuera del sistema.

Observe que todos los miembros de la población potencial alternan entre estar *adentro* y *afuera* del sistema de colas. Entonces, el análogo del modelo *M/M/s* que se ajusta a esta situación supone que el *tiempo afuera* de *cada* miembro (esto es, el tiempo que pasa desde que deja el sistema hasta que regresa) tiene una *distribución exponencial* con parámetro λ. Cuando n miembros están *adentro* y, por supuesto, $N - n$ miembros están *afuera*, la distribución de probabilidad actual del tiempo que *falta* para la próxima llegada al sistema es la distribución del *mínimo* de los *tiempos restantes afuera* de esos $N - n$ miembros. Las propiedades 2 y 3 de la distribución exponencial implican que esta distribución debe ser exponencial con parámetro $\lambda_n = (N - n)\lambda$. Así, el modelo es sólo un caso especial del proceso de nacimiento y muerte que tiene el diagrama de tasas que se presenta en la figura 17.7.

Como $\lambda_n = 0$ para $n = N$, cualquier sistema que se ajuste a este modelo alcanzará en algún momento la condición de estado estable. Los resultados disponibles se resumen de la siguiente manera:

Resultados para el caso de un solo servidor ($s = 1$). Cuando $s = 1$, los factores C_n de la sección 17.5 se reducen a

$$C_n = \begin{cases} N(N-1)\cdots(N-n+1)\left(\dfrac{\lambda}{\mu}\right)^n = \dfrac{N!}{(N-n)!}\left(\dfrac{\lambda}{\mu}\right)^n & \text{para } n \le N \\ 0 & \text{para } n > N, \end{cases}$$

para este modelo. Entonces, si se usa de nuevo la convención de que $n! = 1$ cuando $n = 0$,

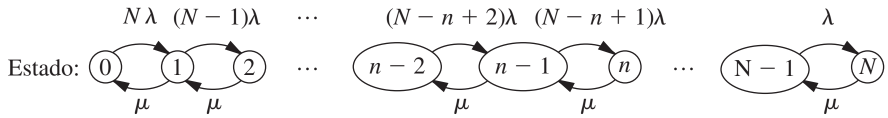
$$P_0 = 1 \Big/ \sum_{n=0}^{N} \left[\frac{N!}{(N-n)!}\left(\frac{\lambda}{\mu}\right)^n\right];$$

$$P_n = \frac{N!}{(N-n)!}\left(\frac{\lambda}{\mu}\right)^n P_0, \qquad \text{si } n = 1, 2, \ldots, N;$$

$$L_q = \sum_{n=1}^{N} (n-1)P_n,$$

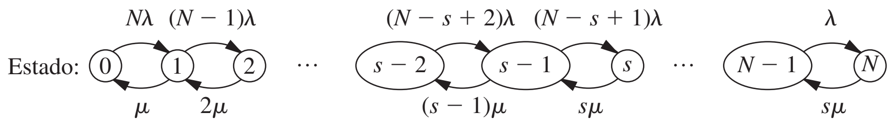
a) *Caso de un solo servidor* ($s = 1$)

$$\lambda_n = \begin{cases}(N-n)\lambda, & \text{para } n = 0, 1, 2, ..., N \\ ?0, & \text{para } n \geq N\end{cases}$$
$$\mu_n = \mu, \quad \text{para } n = 1, 2, ...$$



b) *Caso de varios servidores* ($s > 1$)

$$\lambda_n = \begin{cases}(N-n)\lambda, & \text{para } n = 0, 1, 2, ..., N \\ ?0, & \text{para } n \geq N\end{cases}$$
$$\mu_n = \begin{cases}n\mu, & \text{para } n = 1, 2, ..., s \\ s\mu, & \text{para } n = s, s+1, ...\end{cases}$$



■ **FIGURA 17.7**
Diagrama de tasas de las variaciones de fuente de entrada finita al modelo *M*/*M*/*s*.

que se puede reducir a

$$L_q = N - \frac{\lambda + \mu}{\lambda}(1 - P_0);$$

$$L = \sum_{n=0}^{N} nP_n = L_q + 1 - P_0$$

$$= N - \frac{\mu}{\lambda}(1 - P_0).$$

Por último,

$$W = \frac{L}{\bar{\lambda}} \qquad \text{y} \qquad W_q = \frac{L_q}{\bar{\lambda}},$$

donde

$$\bar{\lambda} = \sum_{n=0}^{\infty} \lambda_n P_n = \sum_{n=0}^{N} (N-n)\lambda P_n = \lambda(N - L).$$

En este punto podría resultar útil referirse en forma retrospectiva al ejemplo del final de la sección 16.8, porque se ajusta por completo a este modelo del caso de un solo servidor. En particular, $N = 2$, $\lambda = 1$ y $\mu = 2$ para ese ejemplo, así que $P_0 = 0.4$, $P_1 = 0.4$, $P_2 = 0.2$, y así sucesivamente.

Resultados para el caso de varios servidores (s . 1). Para $N \geq s > 1$,

$$C_n = \begin{cases} \dfrac{N!}{(N-n)!\,n!}\left(\dfrac{\lambda}{\mu}\right)^n & \text{para } n = 0, 1, 2, \ldots, s \\[2ex] \dfrac{N!}{(N-n)!\,s!\,s^{n-s}}\left(\dfrac{\lambda}{\mu}\right)^n & \text{para } n = s, s+1, \ldots, N \\[2ex] 0 & \text{para } n > N. \end{cases}$$

Entonces,

$$P_n = \begin{cases} \dfrac{N!}{(N-n)!\,n!}\left(\dfrac{\lambda}{\mu}\right)^n P_0 & \text{si } 0 \leq n \leq s \\[2ex] \dfrac{N!}{(N-n)!\,s!\,s^{n-s}}\left(\dfrac{\lambda}{\mu}\right)^n P_0 & \text{si } s \leq n \leq N \\[2ex] 0 & \text{si } n > N, \end{cases}$$

donde

$$P_0 = 1\bigg/\left[\sum_{n=0}^{s-1} \frac{N!}{(N-n)!\,n!}\left(\frac{\lambda}{\mu}\right)^n + \sum_{n=s}^{N} \frac{N!}{(N-n)!s!s^{n-s}}\left(\frac{\lambda}{\mu}\right)^n\right].$$

Por último,

$$L_q = \sum_{n=s}^{N} (n-s)P_n$$

y

$$L = \sum_{n=0}^{s-1} nP_n + L_q + s\left(1 - \sum_{n=0}^{s-1} P_n\right),$$

con lo que después se obtienen W y W_q igual que en el caso de un servidor.

El archivo de Excel de este capítulo contiene una plantilla para realizar todos los cálculos anteriores.

Se dispone de tablas extensas[9] para los resultados numéricos de este modelo, tanto en el caso de uno como de varios servidores.

Se ha demostrado[10] que en ambos casos, las fórmulas anteriores para obtener P_n y P_0 (y por ende L_q, L, W y W_q) *también* se cumplen para una generalización de este modelo. En particular, se puede *eliminar* el supuesto de que los tiempos que los miembros de la población potencial pasan *fuera* del sistema de colas tienen una *distribución exponencial*, aunque esto ponga al modelo fuera del contexto del proceso de nacimiento y muerte. Siempre que estos tiempos tengan distribuciones idénticas con media $1/\lambda$ (y se cumpla el supuesto de tiempos de servicio exponencial), estos tiempos fuera pueden tener *cualquier* distribución de probabilidad.

17.7 MODELOS DE COLAS CON DISTRIBUCIONES NO EXPONENCIALES

Todos los modelos de teoría de colas de la sección anterior (excepto el de una generalización) se basan en el proceso de nacimiento y muerte, lo que hace necesario que tanto los tiempos entre llegadas como los de servicio tengan distribuciones *exponenciales*. Como ya se dijo en la sección 17.4, este tipo de *distribuciones* de probabilidad tiene muchas propiedades convenientes para la teoría de colas, pero sólo en cierto tipo de sistemas de colas proporciona un ajuste razonable. En particular, el supuesto de tiempos entre llegadas exponenciales implica que las llegadas ocurren al azar (proceso de entrada de Poisson), lo cual es una aproximación razonable en muchas situaciones pero *no* cuando las llegadas están programadas o reguladas con todo cuidado. Todavía más, las distribuciones de tiempos de servicio reales con frecuencia se desvían bastante de la forma exponencial, en particular cuando los requerimientos de servicio de los clientes son muy parecidos. Por ello, es importante disponer de otros modelos de colas que usen otras distribuciones de probabilidad.

Desafortunadamente, el análisis matemático de los modelos de colas con distribuciones no exponenciales es mucho más difícil. Sin embargo, se han podido obtener algunos resultados útiles con algunos modelos. El análisis está más allá del nivel de este libro, pero en esta sección se resumirán los modelos y se describirán sus resultados.

Modelo $M/G/1$

Como se dijo en la sección 17.2, el modelo $M/G/1$ supone que el sistema de colas tiene un *servidor* y un *proceso de entradas de Poisson* (tiempos entre llegadas exponenciales) con una tasa media de llegadas *fija* λ. Como siempre, se supone que los clientes tienen tiempos de servicio *independientes* con la misma distribución de probabilidad, pero no se imponen restricciones sobre cuál debe ser esta distribución de tiempos de servicio. En realidad, sólo es necesario conocer (o estimar) la media $1/\mu$ y la variancia σ^2 de esta distribución.

[9] L. G. Peck y R. N. Hazelwood, *Finite Queueing Tables*, Wiley, Nueva York, 1958.

[10] B. D. Bunday y R. E. Scraton, "The G/M/r Machine Interference Model", en *European Journal of Operational Research*, **4:** 399-402, 1980.

Cualquier sistema de líneas de espera de este tipo podrá alcanzar, en algún momento, una condición de estado estable si $\rho = \lambda/\mu < 1$. Los resultados de estado estable disponibles[11] de este modelo general son los siguientes:

$$P_0 = 1 - \rho,$$

$$L_q = \frac{\lambda^2\sigma^2 + \rho^2}{2(1-\rho)},$$

$$L = \rho + L_q,$$

$$W_q = \frac{L_q}{\lambda},$$

$$W = W_q + \frac{1}{\mu}.$$

Si se toma en cuenta la complejidad que representa el análisis de un modelo que permite *cualquier* distribución de tiempos de servicio, es notable que se haya podido obtener una fórmula tan sencilla de L_q. Esta fórmula es uno de los resultados más importantes de la teoría de colas gracias a la facilidad con que se aplica y al predominio de los sistemas *M/G/*1 en la práctica. Esta ecuación de L_q (o su contraparte de W_q) con frecuencia recibe el nombre de **fórmula de Pollaczek-Khintchine**, en honor de dos pioneros del desarrollo de teoría de colas que dedujeron la fórmula de manera independiente a principios de la década de 1930.

Observe que para cualquier tiempo de servicio esperado fijo $1/\mu$, L_q, L, W_q y W se incrementan cuando σ^2 aumenta. Este resultado es importante porque indica que la congruencia del servidor tiene gran trascendencia en el desempeño de la instalación de servicio, no sólo en su velocidad promedio. Este punto esencial se ilustra en la siguiente subsección.

Cuando la distribución de los tiempos de servicio es exponencial, $\sigma^2 = 1/\mu^2$ y los resultados anteriores se reducen a los correspondientes al modelo *M/M/*1 que se presentó al inicio de la sección 17.6.

La flexibilidad total en cuanto a la distribución de los tiempos de servicio que proporciona este modelo es en extremo útil, por lo que es lamentable que no se haya tenido éxito en el desarrollo de resultados análogos en el caso de varios servidores. Ahora bien, se han logrado algunos resultados para más de un servidor en los importantes casos especiales descritos en los dos modelos siguientes. (Se dispone de plantillas de Excel en el archivo de este capítulo para realizar los cálculos del modelo *M/G/*1 y los dos modelos que siguen, cuando $s = 1$.)

Modelo *M/D/s*

Cuando el servicio consiste básicamente en la misma tarea rutinaria que el servidor realiza para todos los clientes, tiende a haber poca variación en el tiempo de servicio que se requiere. Muchas veces, el modelo *M/D/s* proporciona una representación razonable de este tipo de situaciones porque supone que todos los tiempos de servicio son iguales a una *constante* fija (la distribución de tiempos de servicio *degenerada*) y que tiene un proceso de entradas de *Poisson* con tasa media de llegadas fija λ.

Cuando sólo se tiene un servidor, el modelo *M/D/*1 es un caso especial del modelo *M/G/*1, donde $\sigma^2 = 0$, con lo que la *fórmula de Pollaczek-Khintchine* se reduce a

$$L_q = \frac{\rho^2}{2(1-\rho)},$$

donde a partir de este valor de L_q se pueden obtener L, W_q y W como ya se demostró. Observe que el valor de estas L_q y W_q es exactamente igual a *la mitad* que en el caso de tiempos de servicio exponenciales de la sección 17.6 (el modelo *M/M/*l) en el que $\sigma^2 = 1/\mu^2$, y entonces al decrecer σ^2 pueden mejorar mucho las medidas de desempeño de un sistema de colas.

[11] También se dispone de una fórmula recursiva para calcular la distribución de probabilidad del número de clientes en el sistema; vea A. Hordijk y H. C. Tijms: "A Simple Proof of the Equivalence of the Limiting Distribution of the Continuous-Time and the Embedded Process of the Queue Size in the M/G/l Queue", en *Statistica Neerlandica*, **36:**97-100, 1976.

En el caso de la versión de más de un servidor de este modelo ($M/D/s$) se dispone de un método complicado[12] para obtener la distribución de probabilidad de estado estable del número de clientes en el sistema y su media [si se supone que $\rho = \lambda/(s\mu) < 1$]. Existen tabulaciones de estos resultados para muchos casos[13] y también se dispone de gráficas como la de la figura 17.8 para las medias (L).

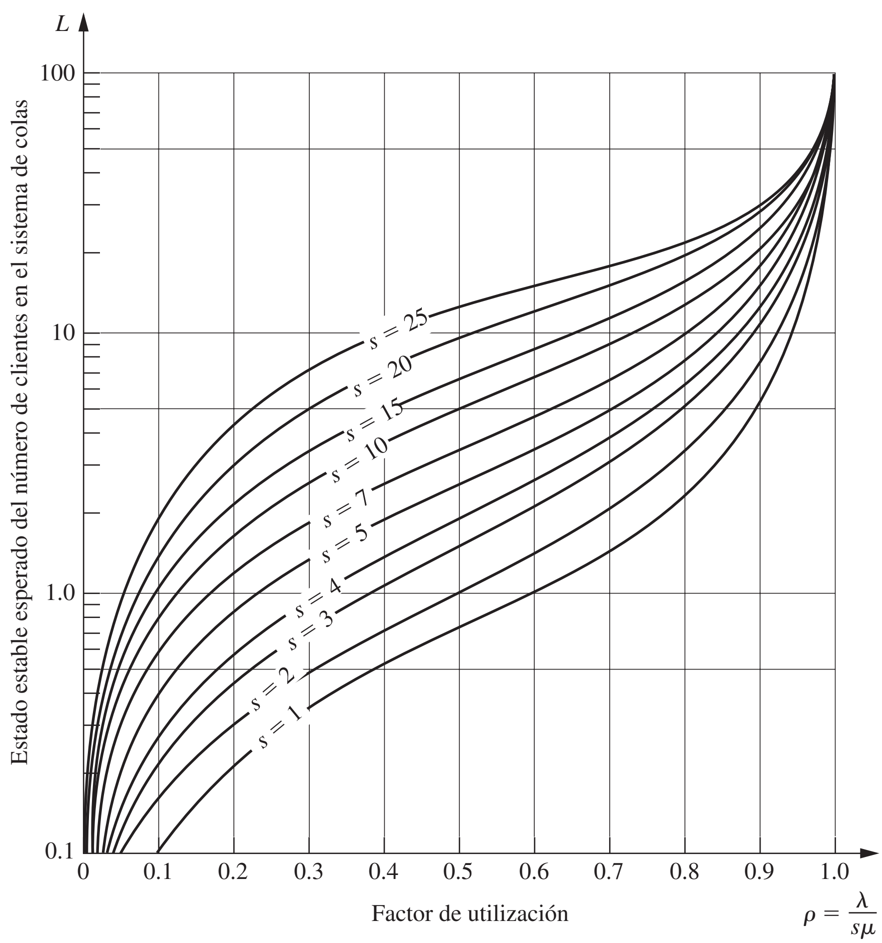
Modelo $M/E_k/s$

El modelo $M/D/s$ supone una variación *cero* en los tiempos de servicio ($\sigma = 0$), mientras que la distribución *exponencial* de tiempos de servicio supone una variación muy grande ($\sigma = 1/\mu$). Entre estos dos casos extremos hay un gran intervalo ($0 < \sigma < 1/\mu$), donde caen la mayor parte de las distribuciones de tiempos de servicio *reales*. Otro tipo de distribución teórica de tiempos de servicio que concuerda con este espacio intermedio es la **distribución de Erlang** (llamada así en honor del fundador de la teoría de colas).

La función de densidad de probabilidad de la distribución de Erlang es

$$f(t) = \frac{(\mu k)^k}{(k-1)!} t^{k-1} e^{-k\mu t}, \qquad \text{para } t \geq 0,$$

FIGURA 17.8
Valores de L del modelo $M/D/s$ (sección 17.7).



[12] Vea N. U. Prabhu: *Queues and Inventories*, Wiley, Nueva York, 1965, pp. 32-34; también vea pp. 286-288 en la referencia seleccionada 5.

[13] F. S. Hillier y O. S. Yu, con D. Avis, L. Fossett, F. Lo y M. Reiman, *Queueing Tables and Graphs*, Elsevier North-Holland, Nueva York, 1981.

donde μ y k son parámetros estrictamente positivos de la distribución y k está restringido a valores enteros. (Excepto por esta restricción de entero y la definición de los parámetros, esta distribución es *idéntica* a la *distribución gamma*.) Su media y desviación estándar son

$$\text{Media} = \frac{1}{\mu}$$

y

$$\text{Desviación estándar} = \frac{\mathbf{1}}{\sqrt{k}}\,\frac{1}{\mu}.$$

En este contexto, k es el parámetro que especifica el grado de variabilidad de los tiempos de servicio con relación a la media. Por lo general, se hace referencia a k como el *parámetro de forma*.

La distribución de Erlang es muy importante en teoría de colas por dos razones. Para describir la primera suponga que $T_1, T_2, \ldots, T_k$ son k variables aleatorias independientes con una distribución *exponencial* idéntica, cuya media es $1/(k\mu)$. Entonces, su suma,

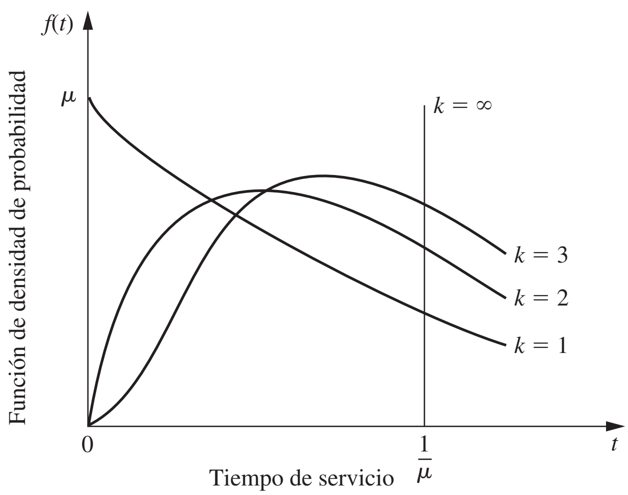
$$T = T_1 + T_2 + \cdots + T_k$$

tiene una distribución de *Erlang* con parámetros μ y k. La presentación de la distribución exponencial en la sección 17.4 sugiere que el tiempo requerido para realizar cierto tipo de tareas podría tener una distribución exponencial. Sin embargo, el servicio total solicitado por un cliente puede incluir una secuencia de k tareas, y no sólo una, realizadas por el servidor. Si las tareas respectivas tienen una distribución exponencial idéntica de su duración, el tiempo total de servicio tendrá una distribución de Erlang; éste sería el caso, por ejemplo, si el servidor debiera realizar la *misma* tarea exponencial k veces para cada cliente.

La distribución de Erlang también es útil debido a que es una gran familia (dos parámetros) de distribuciones que permiten sólo valores no negativos. Así, por lo general se puede obtener una aproximación razonable de la distribución empírica de los tiempos de servicio si se usa una distribución de Erlang. En realidad, tanto la distribución *exponencial* como la *degenerada* (constante) son casos especiales de distribución de Erlang con $k = 1$ y $k = \infty$, respectivamente. Los valores intermedios de k proporcionan distribuciones intermedias con media$= 1/\mu$, moda $= (k-1)/(k\mu)$ y varianza $= 1/(k\mu^2)$, como lo sugiere la figura 17.9. Por lo tanto, después de estimar la media y la varianza de una distribución de servicio empírica, estas fórmulas de la media y la varianza se pueden usar para elegir el valor de k que se ajuste a estas estimaciones de manera más cercana.

Ahora considere el modelo $M/E_k/1$, que es el caso especial del modelo $M/G/1$ donde los tiempos de servicio tienen una distribución de Erlang con parámetro de forma $= k$. Cuando se aplica la

FIGURA 17.9
Familia de distribuciones Erlang con media constante $1/\mu$.



fórmula de Pollaczek-Khintchine con $\sigma^2 = 1/(k\mu^2)$ (y los resultados correspondientes dados por $M/G/1$) se obtiene

$$L_q = \frac{\lambda^2/(k\mu^2) + \rho^2}{2(1-\rho)} = \frac{1+k}{2k}\frac{\lambda^2}{\mu(\mu-\lambda)},$$

$$W_q = \frac{1+k}{2k}\frac{\lambda}{\mu(\mu-\lambda)},$$

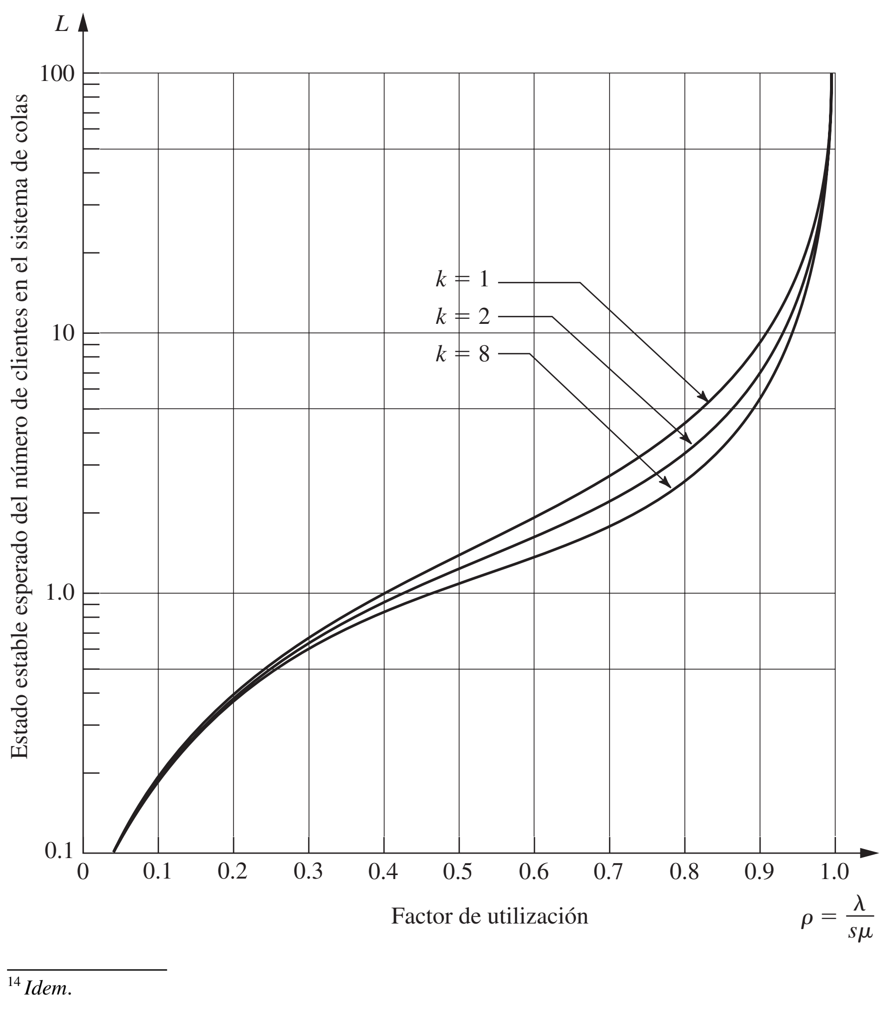
$$W = W_q + \frac{1}{\mu},$$

$$L = \lambda W.$$

Con varios servidores ($M/E_k/s$) se puede aprovechar la relación de la distribución de Erlang con la distribución exponencial que se acaba de describir para formular un proceso de nacimiento y muerte *modificado* (cadena de Markov de parámetro continuo) en términos de las fases del servicio exponencial individual (k por cliente) y no en términos de los clientes. Sin embargo, no ha sido posible derivar una solución general de estado estable [cuando $\rho = \lambda/(s\mu) < 1$] para la distribución de probabilidad del número de clientes en el sistema, como se hizo en la sección 17.5. Más bien se necesitaría una teoría avanzada para resolver en forma numérica los casos individuales. Una vez más, estos resultados se han obtenido y tabulado para casos numéricos.[14] Las medias (L) también se proporcionan en la gráfica de la figura 17.10 para algunos casos donde $s = 2$.

La sección Worked Examples del sitio en internet de este libro incluye **otro ejemplo** en el que se aplica el modelo $M/E_k/s$ para $s = 1$ y $s = 2$ para seleccionar la alternativa menos costosa.

FIGURA 17.10
Valores de L del modelo $M/E_k/2$ (sección 17.7).



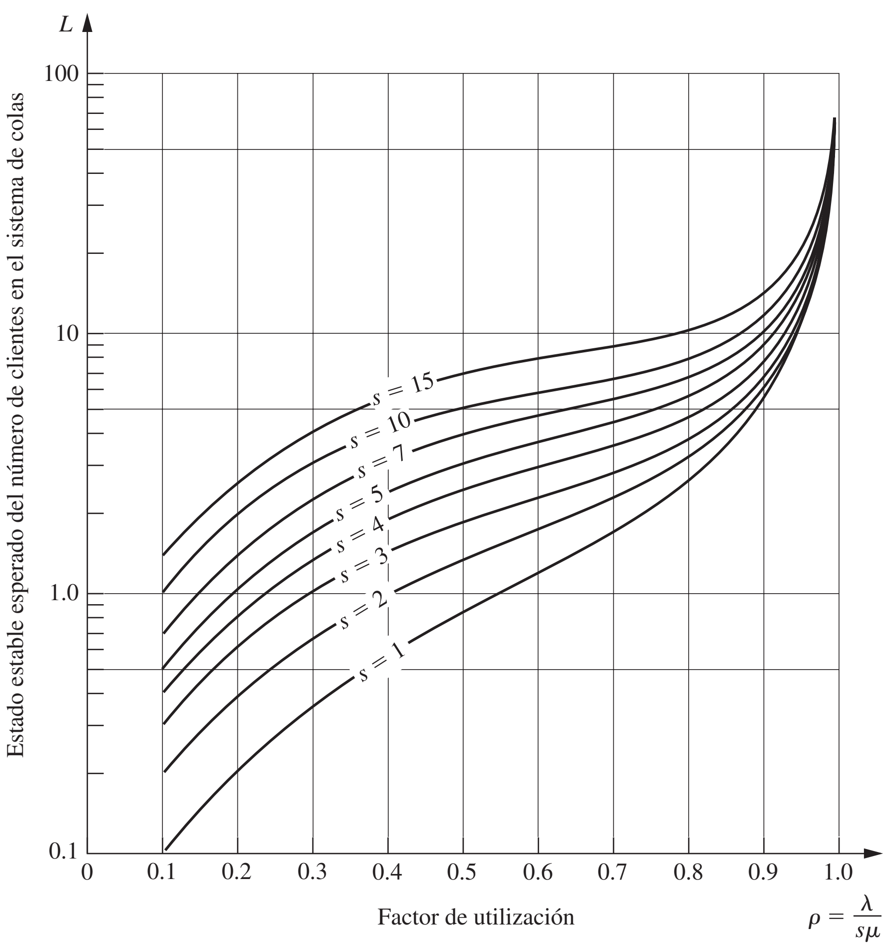
[14] *Idem.*

Modelos sin entradas de Poisson

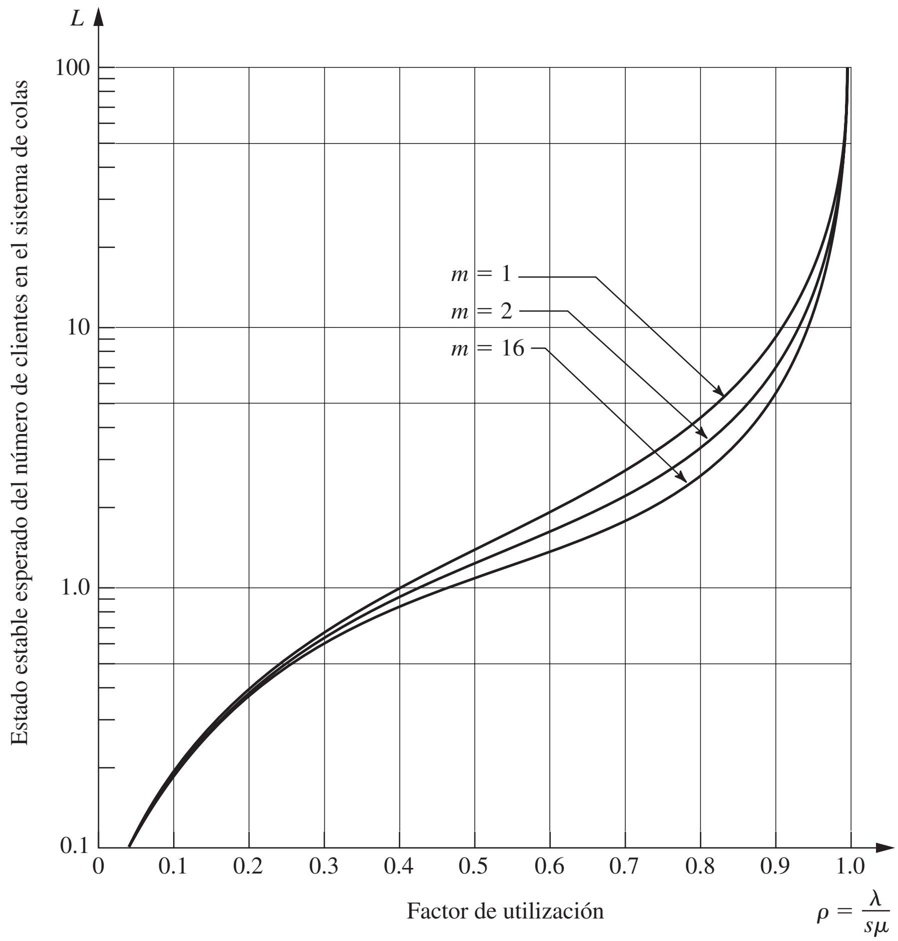
Todos los modelos de colas que se han presentado hasta ahora han supuesto un proceso de entradas de Poisson (tiempos entre llegadas exponenciales). No obstante, este supuesto se viola si las llegadas se programan o regulan de alguna forma que evite que ocurran de manera aleatoria, en cuyo caso se necesita otro modelo.

Se dispone de tres modelos de este tipo siempre que los tiempos de servicio tengan distribución exponencial con un parámetro fijo. Estos modelos se obtienen al *invertir* las distribuciones supuestas de *tiempos entre llegadas* y *tiempos de servicio* de los tres modelos anteriores. El primero de ellos ($GI/M/s$) no impone restricciones sobre el tipo de distribución de los *tiempos entre llegadas*. En este caso se dispone de algunos resultados de estado estable[15] (en especial sobre las distribuciones de tiempos de espera) de las dos versiones del modelo de uno y varios servidores, pero no son ni cercanamente tan convenientes como las expresiones sencillas del modelo $M/G/1$. El segundo modelo nuevo ($D/M/s$) supone que todos los tiempos entre llegadas son iguales a una *constante* fija, que representaría un sistema de colas en el que se *programan* las llegadas a intervalos regulares. El tercer modelo ($E_k/M/s$) supone una distribución de *Erlang* de los tiempos entre llegadas que maneja el espacio intermedio entre llegadas *regulares programadas* (constante) y *completamente aleatorias* (exponencial). En el caso de estos dos últimos modelos, que incluyen los valores de L que se dan en las gráficas de las figuras 17.11 y 17.12, se cuenta con tabulaciones[16] de resultados que se obtuvieron con base en cálculos extensos.

■ FIGURA 17.11
Valores de L del modelo $D/M/s$ (sección 17.7).



[15] Por ejemplo, vea pp. 248-260 de la referencia seleccionada 5.
[16] Hillier y Yu, *op. cit*.



■ **FIGURA 17.12**
Valores de L del modelo $E_k/M/2$ (sección 17.7).

Si ni los tiempos entre llegadas ni el tiempo de servicio de un sistema de colas tienen distribución exponencial, entonces existen tres modelos de colas adicionales para los que también se tienen resultados.[17] Uno de estos modelos ($E_m/E_k/s$) supone una distribución de Erlang de ambos tiempos. Los otros dos ($E_k/D/s$ y $D/E_k/s$) suponen que uno de estos tiempos tiene una distribución de Erlang y el otro es igual a una constante fija.

Otros modelos

Aunque en esta sección se ha visto un gran número de modelos de colas con distribuciones no exponenciales, la lista está muy lejos de agotarse. Por ejemplo, otra distribución que se usa a veces ya sea para los tiempos entre llegadas o para los de servicio es la **distribución hiperexponencial**. La característica esencial de esta distribución es que, si bien sólo permite valores no negativos, su desviación estándar σ es más grande que su media $1/\mu$. Esta característica es opuesta a la distribución de Erlang, donde $\sigma < 1/\mu$ en todos los casos excepto cuando $k = 1$ (distribución exponencial), que tiene $\sigma = 1/\mu$. Para ilustrar una situación representativa en la que puede ocurrir que $\sigma > 1/\mu$, suponga que el servicio en cuestión es la reparación de alguna máquina o vehículo. Si muchas de las reparaciones resultan rutinarias (tiempos de servicio pequeños) pero en ocasiones se requiere una reparación general extensa (con tiempos de servicio grandes), entonces la desviación estándar de los tiempos de servicio tiende a ser grande respecto de la media, en cuyo caso se puede usar la distribución hiperexponencial para representarlos. En particular, esta distribución supone que existen probabilidades fijas, p y $(1 - p)$, para la clase de reparación que ocurrirá, y que el tiempo

[17] *Idem.*

que se requiere para cada clase tiene una distribución exponencial, pero que los parámetros de estas dos distribuciones exponenciales son diferentes. (En general, la distribución hiperexponencial es una composición de este tipo de dos o más distribuciones exponenciales.)

Otra familia de distribuciones que comienza a ser de uso general consiste en **distribuciones tipo fase** (algunas de ellas se llaman *distribuciones erlangianas generalizadas*). Estas distribuciones se obtienen mediante el desglose del tiempo total en cierto número de fases, cada una con distribución exponencial, donde los parámetros de estas distribuciones exponenciales pueden ser diferentes y las fases pueden ser ya sea en serie o en paralelo (o ambos). Un grupo de fases *en paralelo* significa que el proceso elige al azar *una* de las fases cada vez, de acuerdo con probabilidades especificadas. En realidad, este enfoque es la forma en que se deriva la distribución hiperexponencial, por lo que esta distribución es un caso especial de las distribuciones tipo fase. Otro caso especial es la distribución de Erlang, que tiene la restricción de que todas sus k fases están en serie y de que tienen el *mismo* parámetro de sus distribuciones exponenciales. Cuando se eliminan estas restricciones se obtiene mayor flexibilidad en las distribuciones tipo fase para ajustarse a la distribución real de los tiempos entre llegadas o de servicio en los sistemas de colas en estudio. Esta flexibilidad es valiosa en especial cuando, al usar la distribución real en forma directa sobre el modelo, el manejo analítico resulta casi imposible, mientras que la razón de la *media* entre la *desviación estándar* de la distribución real no se acerca a las razones disponibles ($\sqrt{k}$ para $k = 1, 2, \ldots$) de la distribución de Erlang.

Los modelos de colas que usan distribuciones tipo fase todavía se pueden representar mediante una *cadena de Markov de tiempo continuo*, puesto que están construidos a partir de algunas combinaciones de las distribuciones exponenciales. Por lo general, esta cadena de Markov tiene un número infinito de estados, debido a lo cual obtener la distribución de estado estable de los estados del sistema implica resolver un sistema infinito de ecuaciones lineales que tienen una estructura relativamente complicada. La solución de tales sistemas está muy lejos de ser un trabajo rutinario, pero avances teóricos recientes permiten, en algunos casos, resolver estos modelos en forma numérica. Existen tabulaciones extensas[18] de estos resultados de modelos con varias distribuciones tipo fase (que incluyen la distribución hiperexponencial).

17.8 MODELOS DE COLAS CON DISCIPLINA DE PRIORIDADES

En los modelos con disciplina de prioridades, la disciplina de la cola se basa en un *sistema prioritario*. El orden en el que se seleccionan los clientes para darles el servicio se basa en sus prioridades asignadas.

Muchos sistemas reales se ajustan a este tipo de modelos mucho mejor que a otros disponibles. Los trabajos urgentes se hacen antes que otros trabajos y los clientes importantes tienen precedencia sobre otros. Con frecuencia, el uso de modelos con disciplina de prioridades proporciona un refinamiento muy aceptable en comparación con otros modelos de colas más usuales.

Se presentarán dos modelos básicos con disciplina de prioridades. Como ambos se basan en los mismos supuestos, excepto por la naturaleza de las prioridades, primero se describen los modelos juntos y después se resumen sus resultados por separado.

Los modelos

Ambos modelos suponen que existen *N clases de prioridad* (la clase 1 tiene la prioridad más alta y la clase N la más baja) y que siempre que un servidor queda libre para comenzar el servicio de un nuevo cliente, el cliente que se selecciona es el miembro de la clase prioritaria *más alta* representada en la cola y que haya esperado más. En otras palabras, los clientes se seleccionan para comenzar su servicio en el orden de sus clases de prioridad, pero sobre la base de primero en entrar, primero en salir dentro de cada clase prioritaria. Se supone que cada clase prioritaria está sometida a un proceso de entradas de *Poisson* y tiempos de servicio *exponenciales*. El modelo también supone, en cierta manera restrictiva, que el tiempo medio de servicio es el *mismo* para todas las clases prioritarias, pero permite que la tasa media de llegadas difiera entre ellas.

[18] L. P. Seelen, H. C. Tijms y M. H. Van Hoorn: *Tables for Multi-Server Queues*, North-Holland, Amsterdam, 1985.

La distinción entre los dos modelos se basa en que una de las prioridades es *sin interrupción* y la otra *con interrupción*. Cuando se aplica el modelo de **prioridades sin interrupción** no se puede regresar a la cola a un cliente que se encuentra en servicio (interrumpirlo) si entra un cliente de prioridad más alta al sistema de colas. Por lo tanto, una vez que el servidor comienza a atender a un cliente, el servicio debe terminar sin interrupción. El primer modelo supone prioridades sin interrupción.

Cuando se aplica el modelo de **prioridades con interrupción**, el cliente de prioridad más baja que se encuentre en servicio es *interrumpido* (va de regreso a la cola) cada vez que entra un cliente con prioridad más alta al sistema de colas. En esta situación se libera un servidor para que el servicio al recién llegado comience de inmediato. (Cuando un servidor *termina* un servicio, se selecciona el siguiente cliente para comenzar otro servicio como se describió al principio de esta subsección, de manera que, por lo general, un cliente interrumpido regresará a servicio y, después de suficientes intentos, finalmente terminará.) Debido a la propiedad de falta de memoria de la distribución exponencial (vea la sección 17.4), no es necesario preocuparse por definir el punto en el que comienza un servicio cuando regresa un cliente interrumpido; la distribución del tiempo de servicio *restante* es *siempre* la misma. (En cualquier otra distribución de servicio, es importante distinguir entre sistemas *con interrupción y continuación*, en los que el servicio a un cliente que se interrumpió se retoma en el punto de interrupción, y los sistemas *con interrupción y repetición*, donde el servicio debe comenzar de nuevo.) El segundo modelo supone prioridades con interrupción.

En ambos modelos se pasa por alto esta distinción entre clientes en diferentes clases prioritarias. La propiedad 6 de la distribución exponencial (vea la sección 17.4) implica que todos los clientes llegan de acuerdo con un proceso de entradas de Poisson. Más aún, todos los clientes tienen la *misma* distribución exponencial de tiempos de servicio. En consecuencia, los dos modelos en realidad se ajustan al modelo $M/M/s$ que se estudió en la sección 17.6, *excepto* por el orden en que se sirve a los clientes. Entonces, cuando sólo se cuenta el número *total* de clientes en el sistema, se puede aplicar la distribución de estado estable del modelo $M/M/s$ en ambos modelos. Por lo tanto, las fórmulas de L y L_q también sirven, al igual que los resultados del tiempo esperado de espera (de la fórmula de Little), W y W_q, para un cliente elegido en forma aleatoria. Lo que cambia es la *distribución* de los tiempos de espera que se obtuvo en la sección 17.6 bajo el supuesto de una disciplina de primero en llegar, primero en salir. Cuando se aplica la disciplina de prioridades, esta distribución tiene una *varianza* mucho mayor puesto que los tiempos de espera de los clientes de las clases de prioridad alta tienden a ser mucho menores que con la disciplina de primero en entrar, primero en salir, mientras que los tiempos de espera de las clases de prioridad baja tienden a ser mucho mayores. Al seguir el mismo razonamiento, el desglose del número total de clientes en el sistema tiende a poner un peso desproporcionado en las clases prioritarias bajas. Esta condición es el motivo de que se impongan prioridades sobre el sistema de colas. Se desea *mejorar* las *medidas de desempeño* de cada clase prioritaria alta a costa del desempeño de las clases de prioridad baja. Para determinar cuál es la mejora lograda es necesario obtener estas medidas en términos de *tiempo de espera esperado en el sistema* y el *número esperado de clientes en el sistema* de las clases de prioridad individuales. En seguida se obtendrán expresiones de estas medidas en los dos modelos.

Resultados en el modelo con prioridades sin interrupción

Sea W_k el tiempo esperado de espera en el sistema en estado estable (lo cual incluye el tiempo de servicio) de un miembro de la clase de prioridad k. Entonces,

$$W_k = \frac{1}{AB_{k-1}B_k} + \frac{1}{\mu}, \qquad \text{para } k = 1, 2, \ldots, N,$$

donde $$A = s!\frac{s\mu - \lambda}{r^s}\sum_{j=0}^{s-1}\frac{r^j}{j!} + s\mu,$$

$$B_0 = 1,$$

$$B_k = 1 - \frac{\sum_{i=1}^{k}\lambda_i}{s\mu},$$

s = número de servidores,

μ = tasa media de servicio por servidor ocupado,

λ_i = tasa media de llegadas de la clase de prioridad i,

$$\lambda = \sum_{i=1}^{N} \lambda_i,$$

$$r = \frac{\lambda}{\mu}.$$

(Estos resultados suponen que

$$\sum_{i=1}^{k} \lambda_i < s\mu,$$

de manera que la clase de prioridad k puede alcanzar una condición de estado estable.) La *fórmula de Little* se aplica a las clases individuales de prioridad, por lo que L_k, el número esperado de miembros de la clase de prioridad k en el sistema de colas (incluso los que están en servicio), es

$$L_k = \lambda_k W_k, \qquad \text{para } k = 1, 2, \ldots, N.$$

Para determinar el tiempo esperado de espera en la cola (sin incluir el tiempo de servicio) de la clase de prioridad k, sólo se resta $1/\mu$ de W_k; la longitud esperada de la cola correspondiente se obtiene de nuevo si se multiplica por λ_k. En el caso especial en el que $s = 1$, la expresión de A se reduce a $A = \mu^2/\lambda$.

En el OR Courseware se proporciona una plantilla de Excel para realizar los cálculos anteriores.

En la sección Worked Examples del sitio en internet de este libro se da **un ejemplo** que ilustra la aplicación del modelo de prioridades sin interrupción para determinar cuántos tornos debe tener una fábrica cuando los trabajos por realizar caen en tres clases de prioridad.

Variación de un servidor en el modelo de prioridades sin interrupción

El supuesto anterior de que el tiempo de servicio esperado $1/\mu$ es el mismo para todas las clases prioritarias es bastante restrictivo. En la práctica, a veces se viola este supuesto debido a diferencias en los requerimientos de servicio entre las clases de prioridad.

Por fortuna, en el caso especial de un servidor, es posible admitir tiempos esperados de servicio distintos y de todas formas obtener resultados útiles. Sea $1/\mu_k$ la media de la distribución exponencial del tiempo de servicio de la clase de prioridad k, entonces

$$\mu_k = \text{tiempo medio de servicio de la clase de prioridad } k, \quad \text{para } k = 1, 2, \ldots, N.$$

Así, el tiempo esperado de estado estable en el sistema de un miembro de prioridad k es

$$W_k = \frac{a_k}{b_{k-1}b_k} + \frac{1}{\mu_k}, \qquad \text{para } k = 1, 2, \ldots, N,$$

donde $$a_k = \sum_{i=1}^{k} \frac{\lambda_i}{\mu_i^2},$$

$$b_0 = 1,$$

$$b_k = 1 - \sum_{i=1}^{k} \frac{\lambda_i}{\mu_i}.$$

Este resultado es válido mientras

$$\sum_{i=1}^{k} \frac{\lambda_i}{\mu_i} < 1,$$

que permite a la clase con prioridad k alcanzar una condición de estado estable. La fórmula de Little se puede usar como se describió para obtener las otras medidas de desempeño de cada clase prioritaria.

Resultados en el modelo de prioridades con interrupción

En el modelo de prioridades con interrupción, se necesita retomar el supuesto de que el tiempo esperado de servicio es el mismo para todas las clases de prioridad. Con la misma notación que en el modelo anterior, el hecho de poder interrumpir cambia el tiempo esperado *total* en el sistema (incluido el tiempo total de servicio) a

$$W_k = \frac{1/\mu}{B_{k-1}B_k}, \qquad \text{para } k = 1, 2, \ldots, N,$$

en el caso de *un servidor* ($s = 1$). Cuando $s > 1$, la W_k se puede calcular mediante un proceso iterativo que se ilustrará con el ejemplo del Hospital General. La L_k que se acaba de definir todavía satisface la relación

$$L_k = \lambda_k W_k, \qquad \text{para } k = 1, 2, \ldots, N.$$

Los resultados correspondientes de la cola (excluyendo los clientes en servicio) también se pueden obtener a partir de W_k y L_k igual que en el caso de prioridades sin interrupción. Debido a la propiedad de falta de memoria de la distribución exponencial (vea la sección 17.4), las interrupciones no afectan el proceso de servicio (ocurrencias de terminaciones de servicio) de ninguna manera. El tiempo total esperado de servicio de cualquier cliente es $1/\mu$.

El archivo de Excel de este capítulo incluye una plantilla para calcular las medidas de desempeño anteriores en el caso de un solo servidor.

Ejemplo del Hospital General con prioridades

En el problema de la sala de urgencias del Hospital General, el ingeniero administrador observó que los pacientes no son atendidos sobre la base de primero en llegar, primero en salir. En su lugar, parece que la enfermera que realiza las admisiones divide a los pacientes en tres grandes categorías: 1) casos *críticos*, en los que el tratamiento inmediato es vital para la supervivencia; 2) casos *serios*, en los que un tratamiento rápido es importante para prevenir mayor daño, y 3) casos *estables*, en los que el tratamiento puede retrasarse sin consecuencias médicas adversas. Entonces, se atiende a los pacientes en este orden de prioridad, esto es, los pacientes de la misma categoría por lo general son atendidos según la regla de primero en llegar, primero en salir. Un médico interrumpe el tratamiento de un paciente si llega un caso nuevo de una categoría de prioridad más alta. Alrededor de 10% de los pacientes caen en la primera categoría, 30% en la segunda y 60% en la tercera. Debido a que los enfermos más graves se internan en el hospital después de recibir los primeros auxilios, el tiempo promedio de tratamiento por un médico en la sala de urgencias en realidad no difiere mucho entre estas categorías.

El ingeniero ha decidido emplear el modelo de colas con disciplina de prioridades como una representación razonable de este sistema de colas, en el que las tres categorías de pacientes constituyen las tres clases de prioridad del modelo. Como el tratamiento se interrumpe debido a la llegada de un caso de prioridad más alta, la *versión de prioridades con interrupción* es la apropiada. Dados los datos que se tenían desde antes ($\mu = 3$ y $\lambda = 2$), los porcentajes anteriores conducen a $\lambda_1 = 0.2$, $\lambda_2 = 0.6$ y $\lambda_3 = 1.2$. La tabla 17.3 proporciona los tiempos de espera esperados en la cola que resultan (*excluyendo* los tiempos de servicio) para las respectivas clases de prioridad[19] cuando hay un médico ($s = 1$) o dos médicos ($s = 2$) de guardia. (En dicha tabla también se presentan los resultados correspondientes a la versión de prioridades sin interrupción para mostrar el efecto de las interrupciones.)

Obtención de resultados de prioridades con interrupción. Estos resultados de prioridades con interrupción con $s = 2$ se obtuvieron de la manera siguiente. Como los tiempos de espera de los clientes de la clase de prioridad 1 no se ven afectados por la presencia de clientes de clases de prioridad más baja, W_1 será la misma que para cualesquiera otros valores de λ_2 y λ_3, incluyendo

[19] Observe que estos tiempos esperados ya no se pueden interpretar como el tiempo esperado antes de que comience el tratamiento cuando $k > 1$, porque el tratamiento puede ser interrumpido por lo menos una vez, lo que causa un tiempo de espera adicional antes de terminar el servicio.

TABLA 17.3 Resultados de estado estable del modelo de disciplina de prioridades del problema del Hospital General

	Prioridades con interrupción		Prioridades sin interrupción	
	$s = 1$	$s = 2$	$s = 1$	$s = 2$
A	—	—	4.5	36
B_1	0.933	—	0.933	0.967
B_2	0.733	—	0.733	0.867
B_3	0.333	—	0.333	0.667
$W_1 - \frac{1}{\mu}$	0.024 horas	0.00037 horas	0.238 horas	0.029 horas
$W_2 - \frac{1}{\mu}$	0.154 horas	0.00793 horas	0.325 horas	0.033 horas
$W_3 - \frac{1}{\mu}$	1.033 horas	0.06542 horas	0.889 horas	0.048 horas

$\lambda_2 = 0$ y $\lambda_3 = 0$. Entonces, W_1 debe ser igual a W en el modelo correspondiente de *una clase* (modelo *M/M/s* de la sección 17.6) con $s = 2$, $\mu = 3$ y $\lambda = \lambda_1 = 0.2$, que conduce a

$$W_1 = W = 0.33370 \text{ horas}, \qquad \text{para } \lambda = 0.2$$

así

$$W_1 - \frac{1}{\mu} = 0.33370 - 0.33333 = 0.00037 \text{ horas}.$$

Ahora considere las dos primeras clases de prioridad. De nuevo observe que los clientes de las clases de prioridad más baja (en este caso sólo la clase 3) no afectan a los clientes de estas dos clases en nada y por lo tanto se pueden pasar por alto en el análisis. Sea $\overline{W}_{1-2}$ tiempo esperado de espera en el sistema (incluido el tiempo de servicio) de una *llegada aleatoria* en *cualquiera* de estas dos clases, de manera que la probabilidad de que esta llegada pertenezca a la clase 1 es $\lambda_1/(\lambda_1 + \lambda_2) = \frac{1}{4}$ y de que pertenezca a la clase 2 es $\lambda_2/(\lambda_1 + \lambda_2) = \frac{3}{4}$. Por lo tanto,

$$\overline{W}_{1-2} = \frac{1}{4}W_1 + \frac{3}{4}W_2.$$

Lo que es más, como el tiempo de espera *esperado* es el mismo para *cualquier* disciplina de la cola, $\overline{W}_{1-2}$ también debe ser igual a W para el modelo *M/M/s* de la sección 17.6, con $s = 2$, $\mu = 3$ y $\lambda = \lambda_1 + \lambda_2 = 0.8$, lo que conduce a

$$\overline{W}_{1-2} = W = 0.33937 \text{ horas}, \qquad \text{para } \lambda = 0.8.$$

Cuando se combinan estos hechos se obtiene

$$W_2 = \frac{4}{3}\left[0.33937 - \frac{1}{4}(0.33370)\right] = 0.34126 \text{ horas}.$$

$$\left(W_2 - \frac{1}{\mu} = 0.00793 \text{ horas.}\right)$$

Por último, sea $\overline{W}_{1-3}$ el tiempo de espera esperado en el sistema (incluido el tiempo de servicio) de una *llegada aleatoria* de *cualquiera* de las tres clases de prioridad, tal que las probabilidades respectivas de que pertenezca a la clase 1, 2 o 3 son 0.1, 0.3 y 0.6. Entonces,

$$\overline{W}_{1-3} = 0.1W_1 + 0.3W_2 + 0.6W_3.$$

Más aún, $\overline{W}_{1-3}$ también debe ser igual a la W del modelo *M/M/s* en la sección 17.6, con $s = 2$, $\mu = 3$ y $\lambda = \lambda_1 + \lambda_2 + \lambda_3 = 2$, de manera que (de la tabla 17.2)

$$\overline{W}_{1-3} = W = 0.375 \text{ horas}, \qquad \text{para } \lambda = 2.$$

En consecuencia,

$$\begin{aligned} W_3 &= \frac{1}{0.6}\,[0.375 - 0.1(0.33370) - 0.3(0.34126)] \\ &= 0.39875 \text{ horas.} \end{aligned}$$

$$\left(W_3 - \frac{1}{\mu} = 0.06542 \text{ horas.}\right)$$

También se podrían usar, de la misma forma, los resultados correspondientes de W_q del modelo $M/M/s$ de la sección 17.6 para deducir de manera directa las cantidades $W_q - 1/\mu$.

Conclusiones. Cuando $s = 1$, los valores $W_k - 1/\mu$ de la tabla 17.3 en el caso de prioridades con interrupción indican que, si se proporciona un solo médico, puede ocurrir que algunos casos críticos esperen alrededor de un minuto y medio (0.024 horas) en promedio, los casos serios esperarían más de 9 minutos y los casos estables más de 1 hora. (Compare estos resultados con la espera promedio $W_k - \frac{2}{3}$ hora, para todos los pacientes, que se obtuvo en la tabla 17.2 con la disciplina de primero en entrar, primero en salir.) Ahora bien, estos valores representan *esperanzas estadísticas*, por lo que algunos pacientes tendrán que esperar bastante más que el promedio de su clase de prioridad. Esta espera no sería tolerable para los casos críticos y serios, en los que unos minutos pueden ser vitales. Por el contrario, los resultados con $s = 2$ en la tabla 17.3 (prioridades con interrupción) indican que si se agrega un segundo médico, casi se eliminaría la espera de todos los pacientes menos los estables. El ingeniero recomendó que el próximo año se asignaran dos médicos a la sala de urgencias durante las primeras horas de la tarde. El consejo directivo del Hospital General adoptó esta recomendación y, al mismo tiempo, aumentó el costo por el uso de esas instalaciones.

17.9 REDES DE COLAS

Hasta ahora se han tomado en cuenta nada más los sistemas de colas que tienen *una* estación de servicio con uno o más servidores, pero, en realidad, los sistemas de colas que se encuentran en los estudios de IO a veces son *redes de colas*, es decir, redes de instalaciones de servicio en las que los clientes solicitan el servicio de algunas o todas ellas. Por ejemplo, las órdenes que se procesan en un taller se deben programar a través de una secuencia de máquinas entre un grupo (instalaciones de servicio). Es necesario, entonces, estudiar toda la red para obtener información sobre el tiempo esperado total, número esperado de clientes en todo el sistema, etcétera.

Debido a la importancia de las redes de colas, hay mucha actividad de investigación en esta área. Sin embargo, es un campo difícil y aquí se dará sólo una breve introducción.

El descubrimiento e implicaciones de un resultado de importancia fundamental para las redes de colas merecen atención especial. Se trata de la *propiedad de equivalencia* del *proceso de entrada* de los clientes y un *proceso de salida* de los que se van, en ciertos sistemas de colas.

> **Propiedad de equivalencia:** suponga que una instalación de servicio tiene s servidores, un proceso de entradas Poisson con parámetro λ y la misma distribución de los tiempos de servicio de cada servidor con parámetro μ (el modelo $M/M/s$), en donde $s\mu > \lambda$. Entonces, la *salida* en estado estable de esta instalación de servicio también es un proceso de Poisson con parámetro λ.

Observe que esta propiedad no tiene supuestos sobre el tipo de disciplina de la cola que se usa. Ya sea primero en entrar, primero en salir, aleatorio o incluso una disciplina de prioridades como la de la sección 17.8, los clientes que ya fueron atendidos dejarán la instalación de servicio de acuerdo con un proceso de Poisson. La implicación esencial de este hecho para las redes de colas es que si estas unidades tienen que pasar a otra estación para continuar su servicio, esta segunda estación *también* tendrá entradas de Poisson. Con una distribución exponencial de los tiempos de servicio la propiedad de equivalencia se cumplirá también para esta instalación que puede proporcionar entradas de Poisson para una tercera instalación, y así sucesivamente. A continuación se presentarán las consecuencias de dos tipos básicos de redes.

Recuadro de aplicación

Por muchas décadas, **General Motors Corporation (GM)** gozó de tener la posición de líder mundial en la industria automotriz antes de que Toyota ocupara dicho lugar. GM tiene operaciones de manufactura en 32 países, emplea a más de 300 000 personas a nivel mundial y genera ganancias anuales cercanas a los 200 000 millones de dólares. Sin embargo, desde finales de los ochenta, cuando la productividad de las plantas de GM ocupaba un lugar muy por debajo de la industria, la posición en el mercado de la compañía se ha deteriorado debido a la creciente competencia extranjera.

Para contrarrestar los efectos de dicha competencia, la alta administración de GM inició hace muchos años un proyecto de IO a largo plazo con el fin de predecir y mejorar la producción en las cientos de líneas de montaje de la compañía en todo el mundo. El objetivo fue incrementar en gran medida la productividad de sus operaciones de manufactura y, por ende, lograr una ventaja estratégica competitiva.

La herramienta analítica más importante que se utilizó en este proyecto ha sido un *modelo de colas complejo* que utiliza un simple modelo de un solo servidor. El modelo consta de una línea de producción de dos estaciones, cada una de las cuales se modela como un sistema de colas de un solo servidor con tiempos constantes entre llegadas y tiempos constantes de servicio con las excepciones siguientes. En ocasiones, el servidor (comúnmente llamado máquina) de cada estación sale de servicio y no vuelve a trabajar sino hasta que se repara. El servidor de la primera estación también se apaga cuando termina un servicio y el búffer entre las estaciones se llena. El servidor de la segunda estación se detiene cuando termina un servicio y no ha recibido todavía un trabajo proveniente de la primera estación.

El siguiente paso en el análisis consiste en extender este modelo de colas a una línea de producción de dos estaciones a un modelo de una línea de producción con cualquier número de estaciones. Este modelo de colas mayor se utiliza después para analizar la forma en que las líneas de producción deben diseñarse a fin de maximizar su producción. (La técnica de *simulación* que se describirá en el capítulo 20 también se utiliza para este propósito en líneas de producción relativamente complejas.)

Esta *aplicación de la teoría de colas* (y simulación), junto con los sistemas de recolección de datos que la soportan, ha dado enormes beneficios a GM. De acuerdo con fuentes imparciales de la industria, sus plantas, que una vez fueron de las menos productivas en la industria, ahora se encuentran entre las más altas. Las mejoras resultantes en cuanto a producción en más de 30 plantas de automóviles y 10 países han representado *una ganancia de* **2 100 millones de dólares** *en ahorros y ganancias comprobados*. Estos dramáticos resultados llevaron a General Motors a ganar en 2005 el primer lugar en la competencia internacional por el Premio Franz Edelman al desempeño en investigación de operaciones y ciencias de la administración.

Fuente: J. M. Alden, L. D. Burns,T. Costy, R. D. Hutton, C. A. Jackson, D. S. Kim, K. A. Kohls, J. H. Owen, M. A. Turnquist y D. J. Vander Veen, “General Motors Increases Its Production Throughput”, en *Interfaces*, **36**(1): 6-25, enero-febrero de 2006. (En el sitio en internet de este libro www.mhhe.com/hillier se proporciona una liga hacia este artículo.)

Colas infinitas en serie

Suponga que todos los clientes deben recibir servicio en *una serie* de m instalaciones, en una secuencia fija. Además, que cada instalación tiene una cola infinita (el número de clientes que acepta no tiene límite), de manera que las instalaciones en serie forman un sistema de *colas infinitas en serie*. Suponga, además, que los clientes llegan a la primera instalación de acuerdo con un proceso de Poisson con parámetro λ y que cada instalación i ($i = 1, 2, \ldots, m$) tiene la misma distribución exponencial de tiempos de servicio con parámetro μ_i de sus s_i servidores, donde $s_i\mu_i > \lambda$. Debido a la propiedad de equivalencia se puede decir que (en condiciones de estado estable) cada instalación de servicio tiene entrada Poisson con parámetro λ. Entonces, se puede usar el modelo elemental *M*/*M*/*s* de la sección 17.6 (o su contraparte con disciplina de prioridades de la sección 17.8) para analizar cada instalación de servicio en forma independiente de las otras.

Al poder usar el modelo *M*/*M*/*s* para obtener las medidas de desempeño de cada instalación independiente, en lugar de analizar la interacción entre las instalaciones, se logra una simplificación enorme. Por ejemplo, la probabilidad de tener n clientes en una instalación en particular está dada por la fórmula de P_n de la sección 17.6 del modelo *M*/*M*/*s*. La *probabilidad conjunta* de n_1 clientes en la instalación 1, n_2 clientes en la instalación 2, etc., es, entonces, el *producto* de las probabilidades individuales obtenidas de esta manera sencilla. En particular, esta probabilidad conjunta se puede expresar como

$$P\{(N_1, N_2, \ldots, N_m) = (n_1, n_2, \ldots, n_m)\} = P_{n_1}P_{n_2}\cdots P_{n_m}.$$

(Esta forma sencilla de solución se llama **solución en forma de producto**.) De manera similar, el tiempo de espera total esperado y el número esperado de clientes en el sistema completo se pueden obtener con sólo sumar las cantidades correspondientes que se obtuvieron de cada instalación.

Desafortunadamente, la propiedad de equivalencia y sus implicaciones no se cumplen en el caso de colas *finitas* que se presentó en la sección 17.6. En realidad, este caso es importante en la práctica, puesto que, con frecuencia, existen limitaciones definidas de la longitud de la cola que admite cada instalación de servicio de una red. Por ejemplo, es común que se proporcione sólo un pequeño espacio para almacenaje en cada instalación (estación de trabajo) de una línea de producción. Para este tipo de sistemas de colas finitas en serie no se dispone de una solución en forma de producto sencilla. En su lugar, se deben analizar las instalaciones en forma conjunta y sólo se han obtenido resultados limitados.

Redes de Jackson

Los sistemas de colas infinitas en serie no son las únicas redes de colas en las que se puede usar el modelo *M/M/s* para analizar cada instalación de servicio de manera independiente. Otro tipo importante de redes con esta propiedad (solución en forma de producto) son las *redes de Jackson*, que reciben este nombre en honor a James R. Jackson, quien primero estableció, hace ya varias décadas, las características de las redes y demostró que se cumple esta propiedad.

Las características de una red de Jackson son las mismas supuestas para el sistema de colas infinitas en serie, excepto que ahora los clientes visitan las instalaciones en diferente orden (y pueden no llegar a todas). Para cada instalación, los clientes que llegan provienen *tanto* de afuera del sistema (de acuerdo con un proceso de Poisson) como de otras instalaciones. A continuación se resumen estas características.

Una **red de Jackson** es un sistema de m instalaciones de servicio en donde la instalación i ($i = 1, 2, \ldots, m$) tiene:

1. Una cola infinita
2. Clientes que llegan de afuera del sistema según un proceso de entrada Poisson con parámetro a_i
3. s_i servidores con distribución exponencial de tiempos de servicio con parámetro μ_i.

Un cliente que deja la instalación i se encamina después a la instalación j ($j = 1, 2, \ldots, m$) con probabilidad p_{ij} o sale del sistema con probabilidad

$$q_i = 1 - \sum_{j=1}^{m} p_{ij}.$$

Cualquier red de este tipo tendrá la siguiente propiedad.

En condiciones de estado estable, cada instalación j ($j = 1, 2, \ldots, m$) de una red de Jackson se comporta como si fuera un sistema de colas *M/M/s independiente* con tasa de llegadas

$$\lambda_j = a_j + \sum_{i=1}^{m} \lambda_i p_{ij},$$

donde $s_j\mu_j > \lambda_j$.

Esta propiedad clave no se puede *demostrar* en forma directa a partir de la propiedad de equivalencia (el razonamiento lleva a caminar en círculos), pero sí proporciona un *fundamento intuitivo*. Este punto de vista intuitivo (no tan correcto en el sentido técnico) es que, en cada instalación i, los procesos de entrada desde las diferentes fuentes (externa y de otras instalaciones) *son procesos de Poisson independientes*, de manera que el proceso de entrada *agregado* es de Poisson con parámetro λ_i (propiedad 6 de la sección 17.4). La propiedad de equivalencia dice entonces que el proceso *de salida agregado* de la instalación i debe ser Poisson con parámetro λ_i. Al desagregar este proceso de salida (otra vez propiedad 6), el proceso de los clientes que salen de la instalación i a la instalación j debe ser Poisson con parámetro $\lambda_i p_{ij}$. Este proceso se convierte en uno de los procesos de *entrada* de Poisson de la instalación j, lo que ayuda a mantener la serie de procesos de Poisson en todo el sistema.

La ecuación anterior para obtener λ_j se basa en el hecho de que λ_i es tanto la *tasa de salida* como la tasa de entrada de todos los clientes que utilizan la instalación i. Como p_{ij} es la proporción de clientes que salen de la instalación i para ir a la instalación j, la tasa a la que estos clientes de la

instalación i llegan a la instalación j es $\lambda_i p_{ij}$. Al sumar estos productos sobre toda i y después agregar esta suma a a_j, se obtiene la *tasa de llegadas total* a la instalación j desde todas las fuentes.

Para calcular λ_j a partir de esta ecuación se requiere conocer las λ_i para $i \neq j$, pero estas λ_i también son incógnitas dadas por las ecuaciones correspondientes. Por lo tanto, el procedimiento es obtener *simultáneamente* $\lambda_1, \lambda_2, \ldots, \lambda_m$ mediante la solución simultánea de todo el sistema de ecuaciones lineales, con λ_j para $j = 1, 2, \ldots, m$. En el IOR Tutorial se incluye una plantilla de Excel para obtener las λ_j de esta manera.

Para ilustrar estos cálculos, considere una red de Jackson con tres estaciones de servicio que tienen los parámetros que se muestran en la tabla 17.4. Al sustituirlos en la fórmula de λ_j para $j = 1, 2, 3$, se obtiene

$$\begin{aligned}
\lambda_1 &= 1 \qquad\qquad + 0.1\lambda_2 + 0.4\lambda_3 \\
\lambda_2 &= 4 + 0.6\lambda_1 \qquad\quad + 0.4\lambda_3 \\
\lambda_3 &= 3 + 0.3\lambda_1 + 0.3\lambda_2.
\end{aligned}$$

(Piense en cada una de las ecuaciones para ver por qué dan la tasa de llegadas total.) La solución simultánea de este sistema es

$$\lambda_1 = 5, \qquad \lambda_2 = 10, \qquad \lambda_3 = 7\frac{1}{2}.$$

Dada esta solución simultánea, se pueden analizar las tres estaciones de servicio de *manera independiente* mediante las fórmulas del modelo *M/M/s* que se presentó en la sección 17.6. Por ejemplo, para obtener la distribución del número de clientes $N_i = n_i$, en la estación i, note que

$$\rho_i = \frac{\lambda_i}{s_i\mu_i} = \begin{cases} \dfrac{1}{2} & \text{para } i = 1 \\[2ex] \dfrac{1}{2} & \text{para } i = 2 \\[2ex] \dfrac{3}{4} & \text{para } i = 3. \end{cases}$$

Al sustituir estos valores (y los parámetros de la tabla 17.4) en la fórmula de P_n se obtiene,

$$P_{n_1} = \frac{1}{2}\left(\frac{1}{2}\right)^{n_1} \qquad \text{para la instalación 1,}$$

$$P_{n_2} = \begin{cases} \dfrac{1}{3} & \text{para } n_2 = 0 \\[2ex] \dfrac{1}{3} & \text{para } n_2 = 1 \\[2ex] \dfrac{1}{3}\left(\dfrac{1}{2}\right)^{n_2-1} & \text{para } n_2 \geq 2 \end{cases} \qquad \text{para la instalación 2,}$$

$$P_{n_3} = \frac{1}{4}\left(\frac{3}{4}\right)^{n_3} \qquad \text{para la instalación 3.}$$

La *probabilidad conjunta* de (n_1, n_2, n_3) está dada por la forma de solución de producto,

$$P\{(N_1, N_2, N_3) = (n_1, n_2, n_3)\} = P_{n_1}P_{n_2}P_{n_3}.$$

■ **TABLA 17.4 Datos del ejemplo de una red de Jackson**

				p_{ij}		
Instalación j	s_j	μ_j	a_j	$i = 1$	$i = 2$	$i = 3$
$j = 1$	1	10	1	0	0.1	0.4
$j = 2$	2	10	4	0.6	0	0.4
$j = 3$	1	10	3	0.3	0.3	0

De manera similar, el número esperado de clientes L_i, en la estación i se puede calcular como en la sección 17.6:

$$L_1 = 1, \qquad L_2 = \frac{4}{3}, \qquad L_3 = 3.$$

El número *total* de clientes esperado en todo el sistema es

$$L = L_1 + L_2 + L_3 = 5\frac{1}{3}.$$

La obtención de W, el tiempo esperado de espera *total* en el sistema (incluido el servicio) de un cliente tiene un pequeño truco. No se puede simplemente sumar los tiempos de espera esperados en cada una de las estaciones porque un cliente no necesariamente llega a cada una de ellas sólo una vez. No obstante, se puede usar la fórmula de Little, donde la tasa de llegadas al sistema λ es la suma de las tasas de llegadas *desde afuera* a las estaciones, $\lambda = a_1 + a_2 + a_3 = 8$. Entonces,

$$W = \frac{L}{a_1 + a_2 + a_3} = \frac{2}{3}.$$

En conclusión debe señalarse que existen otros tipos de redes de colas (más complicados), donde las estaciones de servicio se pueden analizar de manera independiente de las otras. En realidad, encontrar redes de colas con una solución en forma de producto ha sido como el Santo Grial para la investigación en esta área. Las referencias seleccionadas 3 y 12 contienen información adicional.

17.10 APLICACIÓN DE LA TEORÍA DE COLAS

Debido al valor de la información que proporciona la teoría de colas, ésta se usa con amplitud para dirigir el diseño (o rediseño) de sistemas de líneas de espera. A continuación se explicará cómo se aplica para este fin.

Las decisiones más comunes que deben tomarse cuando se diseña un sistema de colas es cuántos servidores se deben proporcionar. Sin embargo, hay otra serie de decisiones que también pueden ser necesitadas. Las posibles decisiones incluyen:

1. Número de servidores en cada instalación de servicio.
2. Eficiencia de los servidores.
3. Número de instalaciones de servicio.
4. Cantidad de espacio para espera en la cola.
5. Algunas prioridades para diferentes categorías de clientes.

De manera típica, las dos consideraciones primordiales cuando se deben tomar estos tipos de decisiones son 1) el costo en el que se incurre al dar el servicio y 2) las consecuencias de hacer que los clientes esperen en el sistema de colas. Si se proporciona demasiada capacidad de servicio se ocasionan costos excesivos. Si se proporciona una cantidad muy limitada se producen esperas excesivas. En este contexto, la meta es encontrar un trueque adecuado entre el costo del servicio y el tamaño de la espera.

Existen dos enfoques básicos para realizar la búsqueda de este equilibrio. El primero es establecer uno o más criterios para lograr un nivel satisfactorio del servicio en términos de cuánta espera sería aceptable. Por ejemplo, un criterio posible podría ser no exceder cierto número de minutos. Otro podría ser que al menos 95% de los clientes no deben esperar más de determinado número de minutos. También se pueden utilizar criterios similares en términos del número esperado de clientes en el sistema (o la distribución de probabilidad de esta cantidad). El criterio también puede establecerse en términos del tiempo de espera o del número de clientes en la cola en lugar de en el sistema. Una vez que se ha seleccionado el criterio o los criterios, por lo general es sencillo utilizar prueba y error para encontrar el diseño menos costoso del sistema de colas que satisface todos los criterios.

El otro criterio básico para buscar la mejor compensación implica evaluar los costos asociados con las consecuencias de hacer esperar a los clientes. Por ejemplo, suponga que el sistema de colas

es un *sistema de servicio interno* (como se describió en la sección 17.3), donde los clientes son los empleados de una empresa comercial. Si se hace que éstos esperen en el sistema de colas, se ocasiona una *pérdida de productividad*, lo que provoca *pérdidas monetarias*. Esta pérdida es el **costo de espera** asociado con el sistema de líneas de espera. Al expresar este costo de espera como una función del tamaño de la espera, el problema de determinar el mejor diseño del sistema de colas se puede definir como la minimización del *costo total* esperado (costo del servicio más costo de espera) por unidad de tiempo.

A continuación se describe este último enfoque para enfrentar el problema de determinar el número óptimo de servidores que deben proporcionarse.

¿Cuántos servidores deben proporcionarse?

Para formular la función objetivo cuando la variable de decisión es el número de servidores s, sea

$E(\text{CT})$ = costo total esperado por unidad de tiempo,

$E(\text{CS})$ = costo de servicio esperado por unidad de tiempo,

$E(\text{CW})$ = costo de espera por unidad de tiempo.

Entonces, el objetivo es seleccionar el número de servidores para

$$\text{Minimizar } E(\text{CT}) = E(\text{CS}) + E(\text{CW}).$$

Cuando el costo de cada uno de los servidores es el mismo, el **costo de servicio** es

$$E(\text{CS}) = C_s s,$$

donde C_s es el costo marginal de un servidor por unidad de tiempo. Para evaluar CW para cualquier valor de s, observe que $L = \lambda W$ proporciona el tamaño de la espera total en el sistema de colas por unidad de tiempo. Por lo tanto, cuando el costo de espera es proporcional al tamaño de la espera, este costo puede expresarse como

$$E(\text{CW}) = C_w L,$$

donde C_w es el costo de espera por unidad de tiempo de cada cliente en el sistema de colas. En consecuencia, después de estimar las constantes, C_s y C_w, la meta es elegir el valor de s para

$$\text{Minimizar } E(\text{CT}) = C_s s + C_w L.$$

Al elegir el modelo de colas que se ajusta al sistema de colas, el valor de L se puede obtener a partir de varios valores de s. Al aumentar s, disminuye L, al principio con rapidez y después más lentamente.

En la figura 17.13 se presenta la forma general de las curvas de $E(\text{CS})$, $E(\text{CW})$ y $E(\text{CT})$ contra el número de servidores s. (Para una mejor conceptualización, se han dibujado estas curvas como continuas a pesar de que los únicos valores factibles de s son $s = 1, 2, \ldots$) Al calcular $E(\text{CT})$ para valores consecutivos de s hasta que $E(\text{CT})$ deja de disminuir y en su lugar comienza a crecer, es evidente la forma de encontrar el número de servidores que minimiza el costo total. El siguiente ejemplo ilustra este proceso.

Ejemplo

El taller Acme Machine tiene un depósito para las herramientas que necesitan los mecánicos del taller. Los despachadores tienen en su poder las herramientas hasta que llegan los mecánicos y las solicitan. Después las recogen cuando éstos ya no las necesitan. Los supervisores han presentado quejas pues sus mecánicos pasan demasiado tiempo esperando ser atendidos en el depósito de herramientas, por lo que parece que debería haber *más* despachadores. Por otro lado, la administración ejerce presión por reducir gastos en la planta, reducción que conduciría a tener *menos* despachadores. Para resolver estas presiones conflictivas, se ha realizado un estudio de IO para determinar con exactitud cuántos despachadores debe tener el depósito de herramientas.

El depósito de herramientas constituye un sistema de colas, en el cual los despachadores son los servidores y los mecánicos los clientes. Después de recopilar algunos datos sobre los tiempos

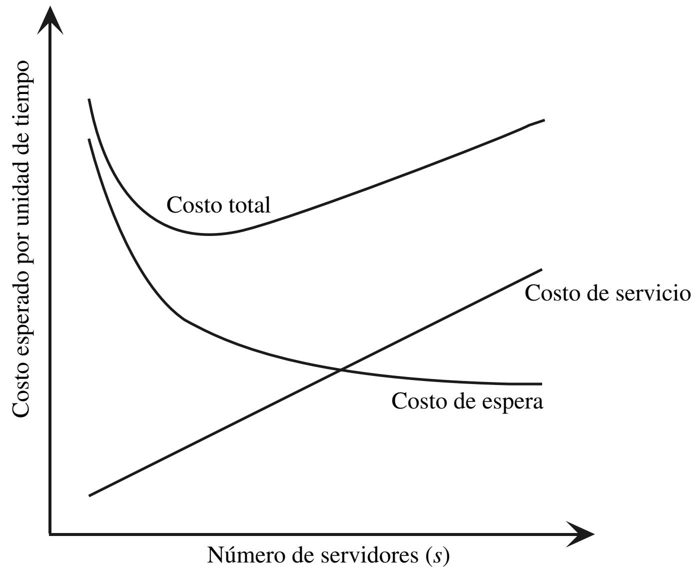


FIGURA 17.13
Forma de las curvas de costo esperado para determinar el número de servidores que se debe proporcionar.

entre llegadas y los tiempos de servicio, el equipo de IO ha concluido que el modelo que se ajusta mejor a este sistema de colas es el *M*/*M*/*s*. Las estimaciones de la tasa media de llegada λ y la tasa media de servicio (por servidor) μ son

$$\lambda = 120 \text{ clientes por hora,}$$
$$\mu = 80 \text{ clientes por hora,}$$

por lo que el factor de utilización de los dos despachadores es

$$\rho = \frac{\lambda}{s\mu} = \frac{120}{2(80)} = 0.75.$$

El costo total para la compañía por cada despachador del depósito de herramientas es de alrededor de \$20 por hora, por lo cual $C_s = \$20$. Mientras un mecánico esté ocupado, el valor de su trabajo para la compañía promedia \$48 por hora, por lo que $C_w = \$48$. Por lo tanto, el equipo de IO necesita determinar el número de servidores (despachadores) s para

$$\text{Minimizar} \quad E(\text{CT}) = \$20\, s + \$48\, L.$$

En el OR Courseware se proporciona una plantilla de Excel para calcular estos costos con el modelo *M*/*M*/*s*. Todo lo que debe hacerse es introducir los datos del modelo junto con el costo de servicio unitario C_s, el costo de espera unitario C_w y el número de servidores s que se desea probar. Entonces la plantilla calcula E(CS), E(CW) y E(CT), operación que se ilustra en la figura 17.14 con $s = 3$ en este ejemplo. Cuando se introduce en forma repetida valores alternativos de s, la plantilla puede revelar cuál valor minimiza E(CT) en cuestión de segundos.

En la tabla 17.5 se muestran los datos que se generarían a partir de esta plantilla al repetir estos cálculos para $s = 1, 2, 3, 4$ y 5. Como el factor de utilización de $s = 1$ es $\rho = 1.5$, un solo despachador sería incapaz de atender a los clientes, por lo que esta opción se debe desechar. Todos los valores más grandes de s son factibles, pero $s = 3$ tiene el costo total esperado más pequeño. Aún más, $s = 3$ reduciría el costo total esperado actual de $s = 2$ en \$61 por hora. Por lo tanto, a pesar de la orden actual de la administración de reducir gastos (lo que incluye el costo de los despachadores en el depósito de herramientas), el equipo de IO recomienda que se agregue un tercer despachador. Observe que esta recomendación reduciría el factor de utilización de los despachadores de un ya modesto 0.75 hasta 0.5. Sin embargo, debido al gran aumento de la productividad de los mecánicos (que son mucho más caros que los despachadores) lograda gracias a la reducción del tiempo que pasan en espera en el depósito de herramientas, la administración adopta la recomendación.

Otros temas

En el capítulo 26 del sitio en internet de este libro se expande de manera considerable la aplicación de la teoría de colas, lo que incluye temas que no se consideraron en este capítulo.

Por ejemplo, el análisis que se muestra en la figura 17.14 y en la tabla 17.5 supone que el tiempo de espera es proporcional a la cantidad de tiempo en espera, pero algunas veces éste no es el caso. Si una compañía tiene a uno o dos de sus empleados en su sistema de colas, ello puede no ser muy serio en términos de la pérdida de productividad porque otros pueden ser capaces de manejar todo el trabajo productivo disponible. Sin embargo, si existen más empleados en el sistema de colas se puede provocar un mayor incremento de pérdida de productividad y la resultante reducción de las ganancias, por lo que el costo de espera se convierte en una función no lineal del número de empleados en el sistema. De manera similar, las consecuencias para un sistema de servicio comercial de hacer esperar a sus clientes pueden ser mínimas en el caso de esperas cortas pero mucho más serias cuando las esperas se prolongan. En este caso, el costo de la espera se convierte en una función no lineal del tiempo de espera. En la sección 26.3 se describe la formulación de las funciones de costos de espera no lineales y el cálculo posterior de E(CW) con dichas funciones.

En la sección 26.4 se expone un modelo de decisión donde las variables de decisión son *tanto* el número de servidores como la tasa media de servicio de los servidores. Un tema interesante que

FIGURA 17.14

La plantilla de Excel en la que se usa análisis económico para seleccionar un número de servidores con el modelo $M/M/s$ se aplica aquí al ejemplo del taller de Acme Machine con $s = 3$.

	A	B	C	D	E	F	G
1	**Análisis económico del ejemplo del taller Acme Machine**						
2							
3			**Datos**				Resultados
4		$\lambda =$	120	(tasa media de llegada)		$L =$	1.736842105
5		$\mu =$	80	(tasa media de servicio)		$L_q =$	0.236842105
6		$s =$	3	(núm. servidores)			
7						$W =$	0.014473684
8		$\Pr(\mathcal{W} > t) =$	0.02581732			$W_q =$	0.001973684
9		si $t =$	0.05				
10						$\rho =$	0.5
11		$\text{Prob}(\mathcal{W}_q > t) =$	0.00058707				
12		si $t =$	0.05			n	P_n
13						0	0.210526316
14		**Análisis económico**				1	0.315789474
15		$C_s =$	$20.00	(costo/servidor/unidad de tiempo)		2	0.236842105
16		$C_w =$	$48.00	(costo de espera/unidad de tiempo)		3	0.118421053
17						4	0.059210526
18		Costo de servicio	$60.00			5	0.029605263
19		Costo de espera	$83.37			6	0.014802632
20		Costo total	$143.37			7	0.007401316

	B	C
18	Costo de servicio	$=C_s*s$
19	Costo de espera	$=C_w*L$
20	Costo total	=CostoDeServicio+CostoDeEspera

Nombre del rango	**Celda**
CostoDeServicio	C18
CostoDeEspera	C19
C_s*s	C15
C_w*L	C16
L	**G4**
s	**C6**
CostoTotal	C20

TABLA 17.5 Cálculo de E(CT) de valores alternativos de s en el ejemplo del taller Acme Machine

s	ρ	L	$E(CS) = C_s s$	$E(CW) = C_w L$	$E(CT) = E(CS) + E(CW)$
1	1.50	∞	\$20	∞	∞
2	0.75	3.43	\$40	\$164.57	\$204.57
3	0.50	1.74	\$60	\$83.37	\$143.37
4	0.375	1.54	\$80	\$74.15	\$154.15
5	0.30	1.51	\$100	\$72.41	\$172.41

surge aquí es si es mejor tener *un servidor rápido* (varias personas que trabajan juntas para servir a cada cliente con rapidez) o *varios servidores lentos* (varias personas que trabajan por separado para servir a diferentes clientes).

También en la sección 26.4 se presenta un modelo de decisión donde las variables de decisión son el número de instalaciones de servicio y el número de servidores por instalación para dar servicio a una población de clientes potenciales. Dada la tasa media de llegada de toda la población, el aumento del número de instalaciones permite la disminución de la tasa media de llegada (carga de trabajo) a cada una de ellas. El número de instalaciones de servicio también afecta cuánto tiempo necesitará cada cliente para trasladarse desde y hacia la instalación más cercana. En este caso, el costo de espera debe ser una función del tiempo total perdido por un cliente mientras espera en una instalación de servicio o se traslada desde y hacia una de ellas. Es por esto que en la sección 26.5 se presentan algunos modelos de tiempo de traslado para determinar el tiempo de viaje esperado de cada cliente.

17.11 CONCLUSIONES

Los sistemas de colas son muy comunes en la sociedad. La adecuación de estos sistemas puede tener un efecto importante sobre la calidad de vida y la productividad.

Para estudiar estos sistemas, la teoría de colas formula modelos matemáticos que representan su operación y luego los utiliza para obtener medidas de desempeño. Este análisis proporciona información vital para diseñar, de manera eficiente, sistemas que logren un balance apropiado entre el costo de prestar el servicio y el asociado con la espera por ese servicio.

En este capítulo se describieron los modelos básicos de teoría de colas de los que se obtienen resultados muy útiles. Se considerarían muchos otros modelos interesantes si el espacio lo permitiera. En realidad, han aparecido en la literatura técnica varios miles de artículos de investigación que formulan y/o analizan modelos de colas, y cada año se publican más.

La *distribución exponencial* tiene un papel fundamental en la teoría de colas para representar la distribución de los tiempos entre llegadas y de servicio, puesto que este supuesto permite representar un sistema de colas como una *cadena de Markov de tiempo continuo*. Por la misma razón, son de gran utilidad las *distribuciones tipo fase* como la *distribución de Erlang*, donde se desglosa el tiempo total en fases individuales que tienen distribuciones exponenciales. Con algunos supuestos adicionales se han obtenido importantes resultados analíticos sólo para un pequeño número de modelos de colas.

Los modelos de disciplina de prioridades son útiles para la situación común en la que se da prioridad a algunas categorías de clientes sobre otras para recibir el servicio.

En otra situación común los clientes deben recibir servicio en distintas estaciones o instalaciones. Los modelos de redes de colas se usan cada vez más en estas situaciones. Ésta es un área especialmente activa en la investigación actual.

Cuando no se dispone de un modelo manejable que proporcione una representación razonable del sistema bajo estudio, un enfoque usual es obtener los datos de desempeño pertinentes mediante el desarrollo de un programa de computadora para simular la operación del sistema. Esta técnica se presenta en el capítulo 20.

En la sección 17.10 se describe brevemente cómo se puede usar la teoría de colas para ayudar a planear en forma adecuada los sistemas de colas y después, en el capítulo 26 (en el sitio en internet de este libro), se hace un estudio más a fondo acerca de este tema.

REFERENCIAS SELECCIONADAS

1. Asmussen, S.: *Applied Probability and Queues*, 2a. ed., Springer, Nueva York, 2003.
2. Balsamo, S., V. de Nitto Personé y R. Onvural: *Analysis of Queuing Networks with Blocking*, Kluwer Academic Publishers (actualmente Springer), Boston, 2001.
3. Chen, H. y D. D. Yao: *Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization*, Springer, Nueva York, 2001.
4. El-Taha, M. y S. Stidham, Jr.: *Sample-Path Analysis of Queueing Systems*, Kluwer Academic Publishers (actualmente Springer), Boston, 1998.
5. Gross, D. y C. M. Harris: *Fundamentals of Queueing Theory*, 3a. ed., Wiley, Nueva York, 1998.
6. Hall, R. W. (ed): *Patient Flow: Reducing Delay in Healthcare Delivery*, Springer, Nueva York, 2006.
7. Hall, R. W.: *Queueing Methods: For Services and Manufacturing*, Prentice Hall, Upper Saddle River, NJ, 1991.
8. Hassin, R. y M. Haviv: *To Queue or Not to Queue; Equilibrium Behavior in Queuing Systems*, Kluwer Academic Publishers (actualmente Springer), Boston, 2008.
9. Hillier, F. S. y M. S. Hillier: *Introduction to Management Science: A Modeling and Case Studies Approach with Spreadsheets*, 3a. ed., McGraw-Hill/Irwin, Burr Ridge, IL, 2008, capítulo 11.
10. Papadopoulos, H. T., C. Heavy y J. Browne: *Queueing Theory in Manufacturing Systems Analysis and Design*, Chapman Hall, Londres, 1993.
11. Prabhu, N. U.: *Foundations of Queueing Theory*, Kluwer Academic Publishers, Boston, 1997.
12. Serfozo, R.: *Introduction to Stochastic Networks*, Springer, Nueva York, 1999.
13. Stidham, S., Jr.: "Analysis, Design, and Control of Queueing Systems", en *Operations Research*, **50**: 197-216, 2002.
14. Tian, N. y Z. G. Zhang: *Vacation Queueing Models: Theory and Applications*, Springer, Nueva York, 2006.

Algunas aplicaciones de la teoría de colas que han ganado premios

(En el sitio en internet de este libro, www.mhhe.com/hillier, se proporciona una liga hacia estos artículos.)

A1. Bleuel, W. H.: "Management Science's Impact on Service Strategy" en *Interfaces*, **5**(1, Parte 2): 4-12, noviembre de 1975.
A2. Brigandi, A. J., D. R. Dargon, M. J. Sheehan y T. Spencer III: "AT&T's Call Processing Simulator (CAPS) Operational Design for Inbound Call Centers", en *Interfaces*, **24**(1): 6-28, enero-febrero de 1994.
A3. Burman, M., S. B. Gershwin y C. Suyematsu: "Hewlett-Packard Uses Operations Research to Improve the Design of a Printer Production Line", en *Interfaces*, **28**(1): 24-36, enero-febrero de 1998.
A4. Quinn, P., B. Andrews y H. Parsons: "Allocating Telecommunications Resources at L. L. Bean, Inc.", en *Interfaces*, **21**(1): 75-91, enero-febrero de 1991.
A5. Ramaswami, V., D. Poole, S. Ahn, S. Byers y A. Kaplan: "Ensuring Access to Emergency Services in the Presence of Long Internet Dial-Up Calls", en *Interfaces*, **35**(5): 411-422, septiembre-octubre de 2005.
A6. Samuelson, D. A.: "Predictive Dialing for Outband Telephone Call Centers", en *Interfaces*, **29**(5): 66-81, septiembre-octubre de 1999.
A7. Swersy, A. J., L. Goldring y E. D. Geyer, Sr.: "Improving Fire Department Productivity: Merging Fire and Emergency Medical Units en New Haven", en *Interfaces*, **23**(1): 109-129, enero-febrero de 1993.
A8. Vandaele, N. J., M. R. Lambrecht, N. De Schuyter y R. Cremmery: "Spicer Off-Highway Products Division-Brugge Improves Its Lead-Time and Scheduling Performance", en *Interfaces*, **30**(1): 83-95, enero-febrero de 2000.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

Ejemplos resueltos:

Ejemplos del capítulo 17

Ejemplos de demostración en el IOR Tutor:

Red de Jackson

Archivos de Excel "Ch. 17—Queueing Theory":

Plantilla para el modelo *M*/*M*/*s*
Plantilla para la variación de cola finita del modelo *M*/*M*/*s*
Plantilla para la variación de población finita del modelo *M*/*M*/*s*
Plantilla para el modelo *M*/*G*/1
Plantilla para el modelo *M*/*D*/1
Plantilla para el modelo $M/E_k/1$
Plantilla para el modelo de prioridades sin interrupción
Plantilla para el modelo de prioridades con interrupción
Plantilla para el análisis económico del número de servidores *M*/*M*/*s*

Archivo de LINGO para ejemplos seleccionados "Ch. 17—Queueing Theory"

Glosario del capítulo 17

Vea el apéndice 1 para la documentación del software.

PROBLEMAS[20]

A la izquierda de los siguientes problemas (o de sus partes) una T indica que el uso de una de las plantillas que se mencionan puede ser útil. Un asterisco en el número del problema indica que al final del libro se proporciona al menos una respuesta parcial.

17.2-1.* Considere una peluquería común. Demuestre que es un sistema de colas mediante la descripción de sus componentes.

17.2-2.* Newell y Jeff son dos peluqueros que operan de manera independiente. Tienen dos sillas para clientes que esperan su corte, porque el número de clientes en el sistema varía entre 0 y 4. Para $n = 1, 2, 3, 4$, la probabilidad P_n de que haya exactamente n clientes en el sistema es $P_0 = \frac{1}{16}$, $P_1 = \frac{4}{16}$, $P_2 = \frac{6}{16}$, $P_3 = \frac{4}{16}$. $P_4 = \frac{1}{16}$.

a) Calcule L. ¿Cómo describiría el significado de L a Newell y Jeff?

b) En el caso de cada valor posible del número de clientes en el sistema, especifique cuántos clientes hay en la cola. Después calcule L_q. ¿Cómo describiría el significado de L_q a Newell y Jeff?

c) Determine el número esperado de clientes que estarán siendo atendidos.

d) Dado que llega un promedio de 4 clientes por hora y esperan el corte, determine W y W_q. Describa estas cantidades en términos que Newell y Jeff comprendan.

e) Dado que Newell y Jeff son igual de rápidos para hacer sus trabajos, ¿cuál es la duración esperada de un corte?

17.2-3. La tienda de alimentos Mom-and-Pop's tiene un estacionamiento pequeño con tres espacios reservados para los clientes. Si la tienda está abierta los autos llegan y usan un espacio con una tasa media de 2 por hora. Para $n = 0, 1, 2, 3$, la probabilidad P_n de que haya exactamente n espacios ocupados es $P_0 = 0.1$, $P_1 = 0.2$, $P_2 = 0.4$, $P_3 = 0.3$.

a) Describa la interpretación de este estacionamiento como un sistema de colas. En particular, identifique los clientes y los servidores. ¿Cuál es el servicio que se proporciona? ¿Qué constituye el tiempo de servicio? ¿Cuál es la capacidad de la cola?

b) Determine las medidas de desempeño básicas: L, L_q, W y W_q de este sistema de colas.

c) Use los resultados de *b*) para determinar el tiempo promedio que un auto permanece en el espacio.

17.2-4. Diga si cada afirmación sobre la cola de un sistema de líneas de espera es falsa o verdadera y justifique su respuesta con referencias específicas al capítulo.

a) La cola es donde los consumidores esperan en el sistema hasta que termina su servicio.

b) Los modelos de colas suponen por convención que la cola puede tener sólo un número limitado de clientes.

c) La disciplina de la cola más común es primero en llegar, primero en salir.

[20] Vea también al final del capítulo 26 (en el sitio en internet de este libro) problemas que implican la aplicación de teoría de colas.

17.2-5. El Midtown Bank siempre tiene dos cajeras en servicio. Los clientes llegan a las cajas a una tasa media de 40 por hora. Una cajera requiere en promedio 2 minutos para servir a un cliente. Cuando ambas cajeras están ocupadas, el cliente que llega se une a una cola y espera a que lo atiendan. Por experiencia se sabe que los clientes esperan en la cola un promedio de 1 minuto antes de pasar a la caja.

a) Describa por qué éste es un sistema de líneas de espera.

b) Determine las medidas de desempeño básicas: W_q, W, L y L_q de este sistema. (*Sugerencia*: No se conocen las distribuciones de probabilidad de los tiempos entre llegadas y los tiempos entre servicio de este sistema de colas, por lo que será necesario usar las relaciones entre estas medidas de desempeño para ayudarle a responder la pregunta.)

17.2-6. Explique por qué el factor de utilización ρ del único servidor del sistema debe ser igual a $1 - P_0$, donde P_0 es la probabilidad de tener 0 clientes en el sistema.

17.2-7. Se tienen dos sistemas de colas, Q_1 y Q_2. La tasa media de llegadas de los clientes, la tasa media de servicio por servidor ocupado y el número esperado de clientes en el estado estable de Q_2 duplican los valores correspondientes de Q_1. Sea W_i = el tiempo esperado de espera en el sistema en estado estable de Q_i, para $i = 1, 2$. Determine W_2/W_1.

17.2-8. Considere un sistema de colas con un servidor que tiene *cualquier* distribución de tiempos de servicio y *cualquier* distribución de tiempos entre llegadas (el modelo *GI*/*G*/1). Use sólo las definiciones básicas y las relaciones de la sección 17.2 para verificar las siguientes relaciones generales:

a) $L = L_q + (1 - P_0)$.

b) $L = L_q + \rho$.

c) $P_0 = 1 - \rho$.

17.2-9. Demuestre que

$$L = \sum_{n=0}^{s-1} nP_n + L_q + s\left(1 - \sum_{n=0}^{s-1} P_n\right)$$

Use las definiciones estadísticas de L y L_q en términos de P_n.

17.3-1. Identifique los clientes y los servidores del sistema de colas en cada una de las situaciones siguientes.

a) La caja de salida de un supermercado.

b) Una estación de bomberos.

c) La caseta de pago para cruzar un puente.

d) Un taller de reparación de bicicletas.

e) Un muelle de carga y descarga.

f) Un grupo de máquinas semiautomáticas asignadas a un operador.

g) El equipo de manejo de materiales de una fábrica.

h) Un taller de plomería.

i) Un taller que produce artículos sobre pedido.

j) Un grupo de secretarias.

17.4-1. Suponga que un sistema de colas tiene dos servidores, distribución de tiempos entre llegadas exponencial con media de 2 horas y distribución de tiempos de servicio exponencial con media de 2 horas para cada servidor. Lo que es más, a las 12:00 del día acaba de llegar un cliente.

a) ¿Cuál es la probabilidad de que la siguiente llegada ocurra *i*) antes de la 1:00 p.m., *ii*) entre la 1:00 y las 2:00 p.m., *iii*) después de las 2:00 p.m.?

b) Suponga que no llegan más clientes antes de la 1:00 p.m. Ahora, cuál es la probabilidad de que la siguiente llegada tenga lugar entre la 1:00 y las 2:00 p.m.?

c) ¿Cuál es la probabilidad de que el número de llegadas entre la 1:00 y las 2:00 p.m. sea *i*) 0, *ii*) 1 y *iii*) 2 o más?

d) Si ambos servidores están ocupados a la 1:00 p.m., ¿cuál es la probabilidad de que *ningún* cliente haya completado su servicio *i*) antes de las 2:00 p.m., *ii*) antes de la 1:10 p.m., *iii*) antes de la 1:01 p.m.?

17.4-2.* Los trabajos que deben realizarse en una máquina específica llegan de acuerdo con un proceso de entradas de *Poisson* con tasa media de 2 por hora. Suponga que la máquina se descompone y su reparación tardará 1 hora. ¿Cuál es la probabilidad de que el número de trabajos que lleguen durante este tiempo sea *a*) 0, *b*) 2, *c*) 5 o más?

17.4-3. El tiempo que requiere un mecánico para reparar una máquina tiene una distribución exponencial con media de 4 horas. Sin embargo, una herramienta especial reduciría esta media a 2 horas. Si el mecánico repara una máquina en menos de 2 horas, se le pagan $100; de otra manera se le pagan $80. Determine el aumento esperado en el pago del mecánico si usa esta herramienta especial.

17.4-4. Un sistema de colas de tres servidores tiene un proceso de llegadas controlado que proporciona clientes a tiempo para mantener siempre ocupados a los servidores. Los tiempos de servicio tienen una distribución exponencial con media de 0.5.

Se observa el arranque del sistema con los tres servidores que inician el servicio en el tiempo $t = 0$. La primera terminación ocurre en $t = 1$. Dada esta información, determine el tiempo esperado después de $t = 1$ hasta que ocurre la siguiente terminación de servicio.

17.4-5. Un sistema de colas tiene tres servidores con tiempos de servicio esperados de 30, 20 y 15 minutos. Los tiempos de servicio tienen una distribución exponencial. Cada servidor ha estado ocupado con el cliente actual durante 10 minutos. Determine el tiempo esperado que falta para la siguiente terminación de un servicio.

17.4-6. Considere un sistema de colas con dos tipos de clientes. Los clientes tipo 1 llegan de acuerdo con un proceso de Poisson a una tasa media de 5 por hora. Los clientes tipo 2 también llegan con un proceso de Poisson a una tasa media de 5 por hora. El sistema tiene dos servidores, que atienden a ambos tipos de clientes. Para los dos tipos el tiempo de servicio tiene una distribución exponencial con media de 10 minutos. El servicio es tipo primero en entrar, primero en salir.

a) ¿Cuál es la distribución de probabilidad (y su media) del tiempo entre llegadas consecutivas de clientes de cualquier tipo?

b) Cuando llega un cliente tipo 2, encuentra dos clientes tipo 1 en el proceso de ser servidos y ningún otro cliente en el sistema. ¿Cuál es la distribución de probabilidad (y su media) del tiempo de espera de este cliente tipo 2 en la cola?

17.4-7. Considere un sistema de colas con dos servidores donde todos los tiempos de servicio son independientes e idénticamente distribuidos según una distribución exponencial con media de 10 minutos. Cuando llega un cliente, encuentra que los dos servidores están ocupados y que no hay nadie en espera en la cola.

a) ¿Cuál es la distribución de probabilidad (con su media y desviación estándar) del tiempo de espera de este cliente en la cola?

b) Determine el valor esperado y la desviación estándar del tiempo de espera de este cliente en el sistema.

c) Suponga que el cliente espera en la cola 5 minutos después de su llegada. Dada esta información, ¿de qué manera cambia el valor esperado y la desviación estándar del tiempo de espera en el sistema a partir de las respuestas que obtuvo en el inciso *b*)?

17.4-8. Diga si cada una de las siguientes afirmaciones de los tiempos de servicio con distribución exponencial es falsa o verdadera y justifique su respuesta con referencia al texto específico (cite la página) en el capítulo.

a) El valor esperado y la varianza de los tiempos de servicio son siempre iguales.

b) La distribución exponencial es una buena aproximación de la distribución de los tiempos de servicio cuando cada cliente requiere las mismas operaciones.

c) En una instalación de s servidores, $s > 1$, con exactamente s clientes en el sistema, una nueva llegada tendrá un tiempo de espera exponencial antes de entrar a servicio de $1/\mu$ unidades de tiempo, donde μ es la media de la tasa de servicio de cada servidor ocupado.

17.4-9. En el caso de la propiedad 3 de la distribución exponencial, sean $T_1, T_2, \ldots, T_n$ variables aleatorias exponenciales independientes con parámetros $\alpha_1, \alpha_2, \ldots, \alpha_n$ y sea $U = \text{mín}\{T_1, T_2, \ldots, T_n\}$. Demuestre que la probabilidad de que cierta variable T_j sea la menor de las n variables es

$$P\{T_j = U\} = \alpha_j \Big/ \sum_{i=1}^{n} \alpha_i, \qquad \text{para } j = 1, 2, \ldots, n.$$

(*Sugerencia:* $P\{T_j = U\} = \int_0^\infty P\{T_i > T_j \text{ para toda } i \neq j | T_j = t\} \alpha_j e^{-\alpha_j t} dt$.)

17.5-1. Considere el proceso de nacimiento y muerte con todas las $\mu_n = 2$ ($n = 1, 2, \ldots$), $\lambda_0 = 3$, $\lambda_1 = 2$, $\lambda_2 = 1$ y $\lambda_n = 0$ para $n = 3, 4, \ldots$

a) Construya el diagrama de tasas.

b) Calcule P_0, P_1, P_2, P_3 y P_n para $n = 4, 5, \ldots$

c) Calcule L, L_q, W y W_q.

17.5-2. Considere el proceso de nacimiento y muerte con tres estados posibles (0, 1 y 2), cuyas probabilidades respectivas de estado estable son P_0, P_1 y P_2. Las tasas de nacimiento y muerte se resumen en la siguiente tabla:

Estado	Tasa de nacimiento	Tasa de muerte
0	4	—
1	2	4
2	0	6

a) Construya el diagrama de tasas de este proceso de nacimiento y muerte.

b) Desarrolle las ecuaciones de balance.

c) Resuelva estas ecuaciones para encontrar P_0, P_1 y P_2.

d) Use las fórmulas generales del proceso de nacimiento y muerte para calcular P_0, P_1 y P_2. También calcule L, L_q, W y W_q.

17.5-3. Considere el proceso de nacimiento y muerte con las siguientes tasas medias. Las tasas de nacimiento son $\lambda_0 = 2$, $\lambda_1 = 3$, $\lambda_2 = 2$, $\lambda_3 = 1$ y $\lambda_n = 0$ para $n > 3$. Las tasas de muerte son $\mu_1 = 3$, $\mu_2 = 4$, $\mu_3 = 1$ y $\mu_n = 2$ para $n > 4$.

a) Construya el diagrama de tasas de este proceso.

b) Desarrolle las ecuaciones de balance.

c) Resuelva estas ecuaciones para encontrar la distribución de probabilidad de estado estable $P_0, P_1, \ldots$

d) Utilice las fórmulas generales del proceso de nacimiento y muerte para calcular $P_0, P_1, \ldots$ También calcule L, L_q, W y W_q.

17.5-4. Considere el proceso de nacimiento y muerte con todas las $\lambda_n = 2$ ($n = 0, 1, \ldots$), $\mu_1 = 2$ y $\mu_n = 4$ para $n = 2, 3, \ldots$

a) Construya el diagrama de tasas.

b) Calcule P_0 y P_1. Después dé una expresión general de P_n en términos de P_0 para $n = 2, 3, \ldots$

c) Considere un sistema de colas de dos servidores que se ajusta a este proceso, ¿cuál es la tasa media de llegadas y la tasa media de servicio por servidor que está ocupado?

17.5-5.* Una gasolinera cuenta con una bomba de gasolina. Los automóviles que desean cargar llegan según un proceso de Poisson a una tasa media de 15 por hora. Sin embargo, si la bomba está en operación, los clientes potenciales pueden *desistir* (ir a otra gasolinera). En particular, si hay n autos en ella, la probabilidad de que un cliente potencial que llega desista es $n/3$ para $n = 1, 2, 3$. El tiempo necesario para servir un auto tiene distribución exponencial con media de 4 minutos.

a) Construya el diagrama de tasas del sistema de colas.

b) Desarrolle las ecuaciones de balance.

c) Resuelva estas ecuaciones para encontrar la distribución de probabilidad de estado estable del número de autos en la gasolinera. Verifique que la solución sea la misma que la solución general del proceso de nacimiento y muerte.

d) Encuentre el tiempo de espera esperado (incluido el servicio) de los automóviles que se quedan.

17.5-6. Un trabajador debe mantener dos máquinas en condiciones de operación. El tiempo que trabaja una máquina antes de descomponerse tiene distribución exponencial con media de 10 horas. El tiempo que tarda el empleado para reparar la máquina tiene distribución exponencial con media de 8 horas.

a) Demuestre que este proceso se ajusta al proceso de nacimiento y muerte mediante la definición de los estados, los valores de $\lambda_n = 2$ y μ_n, y la construcción del diagrama de tasas.

b) Calcule las P_n.

c) Calcule L, L_q, W y W_q.

d) Determine el porcentaje de tiempo que el trabajador de mantenimiento está ocupado.

e) Determine el porcentaje de tiempo que trabaja cada una de las máquinas.

f) Consulte el ejemplo casi idéntico de *cadenas de Markov de tiempo continuo* de la sección 16.8. Describa la relación entre la cadena de Markov de tiempo continuo y el proceso de nacimiento y muerte que permite aplicar cualquiera de los dos al mismo problema.

17.5-7. Considere un sistema de colas de un servidor donde los tiempos entre llegadas tienen una distribución exponencial con parámetro λ y los tiempos de servicio tienen una distribución exponencial con parámetro μ. Además, los clientes *desisten* (dejan el sistema de colas sin pasar a servicio) si su tiempo de espera en la cola es demasiado largo. En particular, suponga que el tiempo que cada cliente está dispuesto a esperar en la cola antes de desistir tiene una distribución exponencial con media de $1/\theta$.

a) Construya el diagrama de tasas de este sistema.
b) Desarrolle las ecuaciones de balance.

17.5-8* Un supermercado pequeño tiene una sola caja con un cajero de tiempo completo. Los clientes llegan a la caja de manera "aleatoria" (proceso de entradas de Poisson) con tasa media de 30 por hora. Cuando sólo hay un cliente en la caja, el cajero lo atiende solo, con un tiempo de servicio esperado de 1.5 min, pero el muchacho que ayuda tiene instrucciones fijas de que si hay más de un cliente en la caja ayude al cajero a empacar la mercancía. Esta ayuda reduce el tiempo esperado de servicio a 1 min. En ambos casos, la distribución de estos tiempos de servicio es exponencial.
a) Construya el diagrama de tasas de este sistema.
b) Obtenga la distribución de probabilidad de estado estable del número de clientes en la caja.
c) Obtenga L de este sistema. (*Sugerencia:* Consulte el desarrollo de L del modelo $M/M/1$ al principio de la sección 17.6.) Utilice esta información para determinar L_q, W y W_q.

17.5-9. Un departamento de una empresa tiene una operadora de procesador de textos. Los documentos que se producen en él se entregan para ser procesados de acuerdo con un proceso de Poisson con un tiempo esperado entre llegadas de 30 min. Cuando la operadora tiene sólo un documento que procesar el tiempo esperado de servicio es de 20 minutos. Cuando hay más de un documento, la ayuda de edición reduce este tiempo a 15 min. En ambos casos, los tiempos de servicio tienen distribución exponencial.
a) Construya el diagrama de tasas de este sistema.
b) Encuentre la distribución de estado estable del número de documentos que la operadora ya recibió pero todavía no procesa.
c) Desarrolle L de este sistema. (*Sugerencia*: consulte el desarrollo de L del modelo $M/M/1$ de la sección 17.6.) Use esta información para determinar L_q, W y W_q.

17.5-10. Los clientes llegan a un sistema de colas de acuerdo con un proceso Poisson con tasa media de llegadas de 2 clientes por minuto. Los tiempos de servicio tienen distribución exponencial con media de 1 minuto. Existe un número ilimitado de servidores por lo que los clientes nunca esperan que comience su servicio. Calcule las probabilidades de estado estable de que haya justo un cliente en el sistema.

17.5-11. Suponga que un sistema de colas de un servidor cumple todos los supuestos del proceso de nacimiento y muerte *excepto* que los clientes siempre llegan en *pares*. La tasa media de llegadas es de 2 pares por hora (4 clientes por hora) y la tasa media de servicio (cuando el servidor está ocupado) es de 5 clientes por hora.
a) Construya el diagrama de tasas de este sistema.
b) Desarrolle las ecuaciones de balance.
c) Con propósitos de comparación, dibuje el diagrama de tasas del sistema correspondiente que se ajusta por completo al proceso de nacimiento y muerte, es decir, los clientes llegan *individualmente* a una tasa media de 4 horas.

17.5-12. Considere un sistema de líneas de espera de un solo servidor con cola finita que puede manejar un máximo de 2 clientes *excluyendo* los que están en servicio. El servidor puede proporcionar *servicio en grupo* a 2 clientes a la vez; el tiempo de servicio tiene distribución exponencial con media de 1 unidad de tiempo independiente del número que se atiende. Cuando la cola no está llena, los clientes llegan en forma individual según un proceso de Poisson con tasa media de 1 por unidad de tiempo.
a) Suponga que el servidor *debe* servir a 2 clientes al mismo tiempo. Por lo tanto, aunque el servidor esté desocupado cuando hay sólo 1 cliente en el sistema, debe esperar a que llegue otro antes de comenzar el servicio. Formule el modelo de colas como una cadena de Markov de tiempo continuo, defina los estados y construya el diagrama de tasas. Elabore las ecuaciones de balance; no las resuelva.
b) Ahora suponga que el tamaño del grupo es 2 sólo si hay 2 clientes en el sistema al terminar el servicio anterior. Por ello, si el servidor está libre cuando hay nada más 1 cliente, deberá atenderlo y cualquiera que llegue después deberá esperar en la cola hasta que el servidor termine con este cliente. Formule el modelo de colas que resulta como cadena de Markov de tiempo continuo; defina los estados y construya el diagrama de tasas. Construya las ecuaciones de balance, pero no las resuelva.

17.5-13. Considere un sistema de colas con dos clases de clientes, dos servidores para atenderlos y *sin cola*. Los clientes potenciales de cada clase llegan según un proceso Poisson con tasa media de 10 clientes por hora de la clase 1 y 5 clientes por hora de la clase 2, pero estos clientes se pierden si no pueden entrar a servicio de inmediato.

Cada cliente de la clase 1 que entra al sistema recibirá servicio de cualquiera de los servidores que esté desocupado; los tiempos de servicio tienen distribución exponencial con tasa media de 5 minutos.

Cada cliente de la clase 2 que entra al sistema requiere la *atención simultánea de los dos servidores* (los dos trabajan juntos como un servidor); los tiempos de servicio tienen distribución exponencial con media de 5 minutos. En consecuencia, un cliente de este tipo se pierde a menos que ambos servidores estén libres para servirlos de inmediato.
a) Formule el modelo como una cadena de Markov de tiempo continuo; defina los estados y el diagrama de tasas.
b) Describa cómo puede ajustarse la formulación del inciso *a*) al proceso de nacimiento y muerte.
c) Use los resultados del proceso de nacimiento y muerte para calcular la distribución conjunta de estado estable del número de clientes de cada clase en el sistema.
d) Para cada clase de clientes, ¿cuál es la fracción esperada de llegadas que no pueden entrar al sistema?

17.6-1. Lea el artículo de referencia que describe el estudio de IO que se resume en la Viñeta de aplicación que se presentó en la sección 17.6. Describa de manera breve la forma en que se aplicó la teoría de colas a este estudio. Después, elabore una lista de los beneficios financieros y no financieros que arrojó dicho estudio.

17.6-2.* La compañía 4M tiene un torno como pieza central del trabajo de la planta. Los trabajos llegan según un proceso Poisson con tasa media de 2 por día. El tiempo de procesado de cada trabajo tiene distribución exponencial con media de $\frac{1}{4}$ día. Como los trabajos son grandes, si no están en proceso se guardan en un almacén a cierta distancia de la máquina. Sin embargo, para ahorrar tiempo al traerlos, el gerente de producción propone agregar espacio para 3 trabajos en proceso además del que está en el torno (el resto seguirá almacenándose). Con esta propuesta, ¿qué proporción de tiempo será adecuado el espacio junto al torno para los trabajos en espera?
a) Calcule su respuesta con las fórmulas disponibles.
T ***b***) Use la plantilla de Excel para obtener las probabilidades necesarias para responder a la pregunta.

17.6-3. Los clientes llegan a un sistema de colas de un servidor según un proceso de Poisson con tasa media de 30 por hora. Si el servidor trabaja en forma continua, el número de clientes que puede atender en una hora tiene distribución de Poisson con media de 50. Determine la proporción del tiempo durante el cual nadie espera servicio.

17.6-4. Considere el modelo *M*/*M*/1 con $\lambda < \mu$.

a) Determine la probabilidad de estado estable de que el tiempo real que espera un cliente sea mayor que el tiempo esperado de espera, es decir, $P\{\mathcal{W} > W\}$.

b) Determine la probabilidad de estado estable de que el tiempo real de espera en la cola sea mayor que el tiempo esperado de espera en la cola, esto es, $P\{\mathcal{W}_q > W_q\}$.

17.6-5. Verifique las siguientes relaciones de un sistema de colas *M*/*M*/1:

$$\lambda = \frac{(1 - P_0)^2}{W_q P_0}, \qquad \mu = \frac{1 - P_0}{W_q P_0}.$$

17.6-6. Es necesario determinar cuánto espacio de almacén para material en proceso conviene asignar a un centro de trabajo en una nueva fábrica. Los trabajos llegan de acuerdo con un proceso de Poisson con tasa media de 4 por hora, y el tiempo que se requiere para realizar el proceso necesario tiene distribución exponencial con media de 0.2 horas. Cuando los trabajos que esperan requieren más espacio de almacén que el asignado, el exceso se concentra en un depósito temporal en un lugar menos conveniente. Si cada trabajo requiere un pie cuadrado de suelo en el almacén del centro de trabajo, ¿cuánto espacio se debe proporcionar para acomodar todos los trabajos *a*) 50%, *b*) 90%, *c*) 99% del tiempo? Deduzca una expresión analítica para responder a estas tres preguntas. *Sugerencia:* La suma de una serie geométrica es

$$\sum_{n=0}^{N} x^n = \frac{1 - x^{N+1}}{1 - x}.$$

17.6-7. Considere las siguientes afirmaciones sobre un sistema de colas *M*/*M*/1 y su factor de utilización ρ. Diga si cada una es falsa o verdadera y justifique su respuesta.

a) La probabilidad de que un cliente tenga que esperar antes de ser servido es proporcional a ρ.

b) El número esperado de clientes en el sistema es proporcional a ρ.

c) Si ρ se incrementa de $\rho = 0.9$ a $\rho = 0.99$, el efecto de cualquier otro incremento de ρ sobre L, L_q, W y W_q será relativamente pequeño siempre que $\rho < 1$.

17.6-8. Los clientes llegan a un sistema de colas según un proceso de Poisson con tiempo esperado entre llegadas de 25 minutos. Los tiempos de servicio tienen distribución exponencial con media de 30 minutos.

Diga si cada una de las siguientes afirmaciones sobre el sistema es falsa o verdadera y justifique su respuesta.

a) El servidor definitivamente estará ocupado siempre después de la llegada del primer cliente.

b) La cola crecerá sin límite.

c) Si se agrega un segundo servidor con la misma distribución de servicio, el sistema puede alcanzar una condición de estado estable.

17.6-9. Diga si cada una de las siguientes afirmaciones sobre un sistema de colas *M*/*M*/1 es falsa o verdadera y justifique su respuesta; para ello haga referencia a afirmaciones específicas (cite las páginas) en el capítulo.

a) El tiempo de espera en el sistema tiene distribución exponencial.

b) El tiempo de espera en la cola tiene distribución exponencial.

c) El tiempo de espera condicional en el sistema, dado el número de clientes que hay, tiene distribución de Erlang (gamma).

17.6-10. La tienda de abarrotes Friendly Neighbor tiene una caja de salida con un cajero de tiempo completo. Los clientes llegan al azar con una tasa media de 20 por hora. La distribución del tiempo de servicio es exponencial con media de 2 minutos. En ocasiones, esta situación causa una cola larga y quejas de los clientes. Como no hay lugar para otra caja, el gerente piensa en contratar a otra persona que ayude a empacar los víveres y reduzca el tiempo esperado de servicio a 1.5 minutos, todavía con distribución exponencial.

El gerente quiere que el porcentaje de tiempo en el que hay más de 2 clientes en la caja fuera menor a 25%. También desea que no más de 5% de los clientes tengan que esperar 5 minutos o más antes de iniciar su servicio o 7 minutos o más antes de terminarlo.

a) Use las fórmulas del modelo *M*/*M*/1 para calcular L, L_q, W y W_q, P_0, P_1 y P_2 de la operación actual. ¿Cuál es la probabilidad de que haya más de dos clientes en la caja?

T ***b***) Use la plantilla de Excel para elaborar este modelo, y verifique sus respuestas en *a*). Encuentre la probabilidad de que el tiempo de espera antes de iniciar el servicio exceda de 5 minutos, y la probabilidad de que el tiempo de espera antes de terminar el servicio sea mayor a 7 minutos.

c) Repita el inciso *a*) para la alternativa de ayuda.

d) Repita el inciso *b*) para esta alternativa.

e) ¿Qué enfoque se debe usar para satisfacer el criterio del gerente lo mejor posible?

T **17.6-11.** El aeropuerto internacional de Centerville tiene dos pistas, una sólo para despegues y otra sólo para aterrizajes. Los aviones llegan al espacio de Centerville para pedir instrucciones de aterrizaje según un proceso de Poisson con tasa media de 10 por hora. El tiempo que se requiere para realizar un aterrizaje después de la aprobación tiene distribución exponencial con media de 3 minutos, proceso que debe estar terminado antes de aprobar otro aterrizaje. Los aviones en espera de pista vuelan en círculos.

La administración de aviación federal tiene varios criterios respecto del nivel seguro de congestión de aviones en espera para aterrizar. Estos criterios dependen de varios factores en cada aeropuerto, como el número de pistas disponibles para ello. En el caso de Centerville los criterios son 1) el número promedio de aviones en espera no debe exceder de 1, 2) 95% del tiempo, el número real de aviones en espera no debe exceder de 4, 3) para 99% de los aviones, el tiempo que vuelan en círculos antes de aterrizar no debe exceder de 30 minutos (exceder este tiempo suele requerir el envío del avión a otro aeropuerto para aterrizaje de emergencia antes de que se agote el combustible).

a) Evalúe la precisión con que se cumplen estos criterios.

b) Una línea aérea importante piensa incluir este aeropuerto en sus rutas, lo cual aumentaría la tasa media de llegadas a 15 aviones por hora. Evalúe con qué exactitud se cumplirían los criterios si dicha empresa comenzara a operar en Centerville.

c) Para atraer más negocios [incluso la línea aérea del inciso *b*)], la administración del aeropuerto piensa añadir otra pista de aterrizaje. Se estima que con el tiempo esta medida aumentaría la

tasa media de llegadas a 25 aviones por hora. Evalúe con qué exactitud se cumplirían los criterios si esto ocurriera.

T **17.6-12.** El Security & Trust Bank tiene 4 cajeros para atender a sus clientes, los cuales llegan según un proceso de Poisson con tasa media de 2 por minuto. Sin embargo, el negocio crece y la gerencia pronostica que esta tasa será de 3 por minuto dentro de un año. El tiempo de transacciones entre el cajero y el cliente tiene distribución exponencial con media de 1 minuto.

La gerencia ha establecido las siguientes políticas para lograr un nivel de servicio satisfactorio. El número promedio de clientes que esperan en la cola para iniciar su servicio no debe exceder de 1. Al menos 95% del tiempo, el número de clientes en la cola no debe exceder de 5. En el caso de al menos 95% de los clientes, el tiempo de espera para iniciar el servicio no debe exceder de 5 minutos.

a) Use el modelo *M*/*M*/*s* para determinar si se cumplen estas políticas.
b) Evalúe si se cumplirán estas políticas dentro de un año si no cambia el número de cajeros.
c) Determine cuántos cajeros se necesitarán dentro de un año para cumplir por completo las políticas.

17.6-13. Considere el modelo *M*/*M*/*s*.

T ***a***) Suponga que sólo hay un servidor y que el tiempo esperado de servicio es exactamente de 1 minuto. Compare L de los casos en que la tasa media de llegadas es 0.5, 0.9 y 0.99 clientes por minuto. Haga lo mismo con L_q, W y W_q y $P\{\mathcal{W} > 5\}$. ¿Qué puede concluir sobre el efecto del factor de utilización creciente ρ de valores pequeños (como, $\rho = 0.5$) a valores grandes (como $\rho = 0.9$) y aún más grandes cercanos a 1 (como, $\rho = 0.99$)?
b) Ahora suponga que se cuenta con dos servidores y que el tiempo esperado de servicio es exactamente de 2 minutos. Siga las instrucciones del inciso *a*).

T **17.6-14.** Considere el modelo *M*/*M*/*s* con tasa media de llegadas de 10 clientes por hora y tiempo de servicio esperado de 5 min. Use la plantilla de Excel de este modelo para obtener e imprimir las medidas de desempeño (con $t = 10$ y $t = 0$ para las dos probabilidades de tiempo de espera) cuando el número de servidores es 1, 2, 3, 4 y 5. Entonces, para los siguientes criterios posibles de nivel satisfactorio de servicio (donde la unidad de tiempo es 1 minuto), use los resultados impresos para determinar cuántos servidores se necesitan para satisfacer los criterios.

a) $L_q \le 0.25$
b) $L \le 0.9$
c) $W_q \le 0.1$
d) $W \le 6$
e) $P\{\mathcal{W}_q > 0\} \le 0.01$
f) $P\{\mathcal{W} > 10\} \le 0.2$
g) $\sum_{n=0}^{s} P_n \ge 0.95$

17.6-15. Una estación de gasolina con sólo una bomba emplea la siguiente política: si un cliente tiene que esperar, el precio es de $3.50 por galón; si no tiene que esperar, el precio es de $4.00 por galón. Los clientes llegan según un proceso de Poisson a una tasa media de 20 por hora. Los tiempos de servicio tienen distribución exponencial con media de 2 minutos. Los clientes que llegan siempre esperan hasta que compran la gasolina. Determine el precio esperado de la gasolina por galón.

17.6-16. Un sistema de colas *M*/*M*/1 tiene una tasa media de llegadas λ y tasa media de servicio μ. Un cliente recibe n dólares si cuando llega hay n clientes en el sistema. Determine el costo esperado en dólares por cliente.

17.6-17. En la sección 17.6 se obtuvieron las siguientes ecuaciones del modelo *M*/*M*/1:

$$(1) \quad P\{\mathcal{W} > t\} = \sum_{n=0}^{\infty} P_n P\{S_{n+1} > t\}.$$

$$(2) \quad P\{\mathcal{W} > t\} = e^{-\mu(1-\rho)t}.$$

Demuestre que la ecuación (1) se reduce en forma algebraica a la ecuación (2). (*Sugerencia:* Use diferenciación, álgebra e integración.)

17.6-18. En los siguientes casos, derive W_q de manera directa; desarrolle y reduzca una expresión análoga a la ecuación (1) del problema 17.6-17. (*Sugerencia:* Use el tiempo de espera esperado *condicional* en la cola dado que una llegada aleatoria encuentra n clientes en el sistema.)

a) El modelo *M*/*M*/1
b) El modelo *M*/*M*/*s*

T **17.6-19.** Considere un sistema de colas *M*/*M*/2 con $\lambda = 3$ y $\mu = 2$. Determine la tasa media a la que ocurren las terminaciones de servicio durante los periodos en que no hay clientes en espera en la cola.

T **17.6-20.** Un sistema de colas *M*/*M*/2 tiene $\lambda = 4$ por hora y $\mu = 6$ por hora. Determine la probabilidad de que un cliente que llega espere más de 30 minutos en la cola dado que hay al menos 2 clientes en el sistema.

17.6-21.* En la compañía de seguros Blue Chip Life, las funciones de depósito de cierto producto de inversión se asignan a un empleado y el retiro a otro. Los formularios de depósito llegan al escritorio de Clara a una tasa media de 16 por hora (proceso Poisson). Las formas de retiro llegan al escritorio de Carmen a una tasa media de 14 por hora (proceso Poisson). El tiempo que se requiere para procesar cualquier transacción tiene distribución exponencial con tasa media de 3 minutos. Para reducir el tiempo esperado de espera en el sistema de ambas formas los actuarios recomiendan: 1) capacitar a las dos empleadas para manejar depósitos y retiros y 2) poner los dos tipos de transacciones en una sola cola con acceso a ambas empleadas.

a) Determine el tiempo esperado de espera en el sistema de cada tipo de llegada con el proceso actual. Combine los resultados y calcule el tiempo esperado de espera en el sistema de una llegada de cualquier tipo.

T ***b***) Si se adoptan las recomendaciones, determine el tiempo esperado de espera en el sistema de las transacciones que llegan.

T ***c***) Suponga que, si se adoptan las recomendaciones, el resultado sería un pequeño aumento en el tiempo esperado de proceso. Utilice la plantilla de Excel de *M*/*M*/*s* para determinar, por prueba y error, el tiempo esperado de proceso (dentro de 0.001 horas) que haría que el tiempo esperado de espera en el sistema de una llegada aleatoria fuera en esencia el mismo de los procesos actuales y con las recomendaciones.

17.6-22. La compañía People's Software acaba de abrir un centro de asistencia técnica por teléfono para su nuevo software. Dos técnicos toman las llamadas; el tiempo que se requiere para responder a las preguntas de cada cliente tiene distribución exponencial con media

de 8 minutos. Las llamadas llegan según un proceso de Poisson con tasa media de 10 por hora.

Se espera que el año próximo la tasa media de llamadas disminuya a 5 por hora, por lo que el plan es reducir el número de técnicos a uno.

T ***a***) Suponga que μ seguirá en 7.5 llamadas por hora el siguiente año, y determine L, L_q, W y W_q de ambos sistemas (actual y en un año). En el caso de cada medida de desempeño, ¿qué sistema da el menor valor?

b) Suponga que μ se ajustará cuando sólo haya un técnico. Obtenga por métodos algebraicos el valor de μ que da el mismo valor de W del sistema actual.

c) Repita el inciso *b*) con W_q en lugar de W.

17.6-23. Considere una generalización del modelo *M*/*M*/1 en la que el servidor necesita "entrar en calor" al principio de un periodo de trabajo y sirve al primer cliente del periodo a una tasa más lenta que a los otros. En particular, si un cliente que llega encuentra el servidor desocupado, experimenta un tiempo de servicio con distribución exponencial con parámetro μ_1. Pero si encuentra al servidor ocupado su tiempo de servicio tendrá distribución exponencial con parámetro μ_2, donde $\mu_1 < \mu_2$. Los clientes llegan de acuerdo con un proceso Poisson con tasa media de λ.

a) Formule este modelo como una cadena de Markov de tiempo continuo; defina los estados y construya el diagrama de tasas correspondiente.

b) Desarrolle las ecuaciones de balance.

c) Suponga que se especifican los valores numéricos de μ_1, μ_2 y λ, y que $\lambda < \mu_2$ (de manera que existe una distribución de estado estable). Como este modelo tiene un número infinito de estados, la distribución de estado estable es la solución simultánea de un número infinito de ecuaciones de balance (más la ecuación de la suma de las probabilidades igual a 1). Suponga que no puede obtener una solución analítica y desea usar la computadora para resolver el modelo numérico. Considere que es imposible resolver un número infinito de ecuaciones y describa en forma breve qué se puede hacer con estas ecuaciones para obtener una aproximación de la distribución de estado estable. ¿En qué circunstancias esta aproximación será esencialmente exacta?

d) Dado que se ha obtenido la distribución de estado estable, dé expresiones explícitas para L, L_q, W y W_q.

e) Dada esta distribución de estado estable, desarrolle una expresión de $P\{\mathcal{W} > t\}$ análoga a la ecuación (1) del problema 17.6-17.

17.6-24. Para cada modelo siguiente, escriba las ecuaciones de balance y demuestre que la solución que se dio en la sección 17.6, de la distribución de estado estable del número de clientes en el sistema, las satisface.

a) Modelo *M*/*M*/1.

b) Variación de cola finita de *M*/*M*/1, con $K = 2$.

c) Variación de fuente de entrada finita de *M*/*M*/1, con $N = 2$.

T **17.6-25.** Considere un sistema telefónico con tres líneas. Las llamadas siguen un proceso de Poisson con una tasa media de 6 por hora. La duración de las llamadas tiene distribución exponencial con tasa de media de 15 minutos. Si todas las líneas están ocupadas, las llamadas se ponen en espera hasta que esté disponible una línea.

a) Imprima las medidas de desempeño de la plantilla de Excel de este sistema (con $t = 1$ hora y $t = 0$ de las probabilidades respectivas de tiempo de espera).

b) Utilice el resultado impreso de $P\{\mathcal{W}_q > 0\}$ para identificar la probabilidad de estado estable de que una llamada sea tomada de inmediato (sin quedar en espera). Verifique esta probabilidad con los resultados de P_n.

c) Utilice los resultados impresos para identificar las probabilidades de estado estable del número de llamadas en espera.

d) Imprima las nuevas medidas de desempeño si las llamadas se pierden cuando todas las líneas están ocupadas. Con los resultados identifique la probabilidad de estado estable de que una llamada que llega se pierda.

17.6-26.* Janet planea abrir un pequeño autolavado y debe decidir cuánto espacio debe asignar a los autos que esperan. Ella estima que los clientes llegarán de manera aleatoria (proceso Poisson) a una tasa media de 1 cada 4 minutos, a menos que el área de espera esté llena, en cuyo caso los clientes que llegan llevarán su auto a otra parte. El tiempo total atribuible al lavado de un automóvil tiene distribución exponencial con media de 3 minutos. Compare la fracción de clientes potenciales que se *pierden* por falta de espacio de espera si se proporcionan *a*) 0, *b*) 2 y *c*) 4 espacios (además del lugar de lavado).

17.6-27. Considere la variación de cola finita del modelo *M*/*M*/*s*. Derive la expresión de L_q dada en la sección 17.6 de este modelo.

17.6-28. En el caso de la variación de cola finita del modelo *M*/*M*/1, desarrolle una expresión análoga a la ecuación (1) del problema 17.6-17 de las siguientes probabilidades:

a) $P\{\mathcal{W} > t\}$.

b) $P\{\mathcal{W}_q > t\}$.

[*Sugerencia:* Las llegadas ocurren sólo cuando el sistema no está lleno, por lo que la probabilidad de que una llegada aleatoria encuentre n clientes es $P_n/(1 - P_K)$.]

17.6-29. George planea abrir un negocio de revelado de fotografías que dará atención a los clientes en su automóvil, con una sola ventanilla de servicio que abrirá alrededor de 200 horas al mes en un área comercial muy activa. El espacio de un carril de espera de los automóviles de los clientes está disponible por una renta de $200 al mes por cada lugar de automóvil. George necesita decidir cuántos lugares de automóvil debe proporcionar a sus clientes.

Si se excluye este costo de renta del carril de automóviles en espera, George cree que promediará una ganancia de $4 por cliente servido (nada para un cliente que va a entregar su película y $8 cuando se recogen las fotografías). También estima que los clientes llegarán a una tasa media de 20 por hora, aunque aquellos que encuentren el carril lleno se verán obligados a irse. La mitad de los clientes que encuentren el carril lleno deseará entregar su película, mientras que la otra mitad deseará recoger sus fotografías. La mitad que desea entregar la película hará dicha entrega en alguna otra parte. La otra mitad de los clientes que encuentren lleno el carril no se perderá porque tratará después hasta que pueda entrar para recoger sus fotografías. George supone que el tiempo necesario para servir a un cliente tendrá una distribución exponencial con una media de 2 minutos.

T ***a***) Encuentre L y la tasa media a la cual los clientes se pierden cuando el número de lugares para automóvil es 2, 3, 4 y 5.

b) Calcule W a partir de L en los casos considerados en *a*).

c) Utilice los resultados en *a*) para calcular la disminución de la tasa media a la que los clientes se pierden cuando el número de lugares para automóvil se incrementa de 2 a 3, de 3 a 4 y de 4 a 5. Después calcule el aumento de la ganancia esperada por hora

(sin incluir los costos de renta por el espacio para automóviles) en cada uno de estos tres casos.

d) Compare el incremento de la ganancia esperada que se encontró en *c*) con el costo por hora de rentar cada lugar para automóvil. ¿Qué conclusión obtiene acerca del número de lugares para automóvil que George debe proporcionar?

17.6-30. En Forrester Manufacturing se ha asignado a un técnico el mantenimiento de tres máquinas. La distribución de probabilidad del tiempo de operación de cada máquina antes de descomponerse es exponencial con media de 9 horas. El tiempo de reparación también tiene distribución exponencial con media de 2 horas.

a) ¿Qué modelo de colas se ajusta a este sistema?

T ***b***) Con este modelo encuentre la distribución de probabilidad del número de máquinas descompuestas y su media.

c) Use la media para calcular el tiempo esperado entre la descompostura de una máquina y la terminación de su reparación.

d) ¿Cuál es la fracción esperada de tiempo que el técnico está ocupado?

T ***e***) Como aproximación burda, suponga que la fuente de entrada es infinita y las entradas son de Poisson con una tasa media de 3 cada 9 horas. Compare el resultado del inciso *b*) con el que obtenga con esta aproximación *i*) con el modelo *M*/*M*/*s* y *ii*) con el modelo de cola finita *M*/*M*/*s* con $K = 3$.

T ***f***) Repita *b*) con un segundo técnico que ayuda a reparar una segunda máquina si más de una de ellas requiere reparación.

17.6-31. Reconsidere el proceso de nacimiento y muerte específico descrito en el problema 17.5-1.

a) Identifique un modelo de colas (y el valor de sus parámetros) de la sección 17.6 que se ajuste a este proceso.

T ***b***) Use la plantilla de Excel correspondiente para obtener las respuestas a los incisos *b*) y *c*) del problema 17.5-1.

T **17.6-32.*** La Dolomite Corporation planea una nueva fábrica. Se han asignado máquinas automáticas a un departamento. Se contratará un pequeño número (a determinar) de operarios para darles servicio ocasional (carga, descarga, ajuste, preparación, etc.), y debe decidirse cómo organizarlos. La alternativa 1 es asignar a cada operario sus máquinas. La alternativa 2 es agruparlos para que el operario desocupado dé servicio a la siguiente máquina que lo necesite. La alternativa 3 es combinarlos en una cuadrilla y que juntos atiendan cualquier máquina.

El tiempo de operación (entre la terminación de un servicio y la necesidad de atender una máquina de nuevo) de cada máquina tiene distribución exponencial con media de 150 minutos. El tiempo de servicio tiene distribución exponencial con media de 15 minutos (alternativas 1 y 2) o 15 minutos dividido entre el número de operarios (alternativa 3). Para que el departamento logre la tasa de producción que se requiere, las máquinas deben operar por lo menos 89% del tiempo en promedio.

a) En el caso de la alternativa 1, ¿cuál es el número máximo de máquinas asignadas a un operario para lograr la tasa de producción? ¿Cuál es la utilización resultante de cada operario?

b) En el caso de la alternativa 2, ¿cuál es el número mínimo de operarios necesario para lograr la tasa de producción? ¿Cuál es la utilización de cada operario?

c) En el caso de la alternativa 3, ¿cuál es el tamaño mínimo de la cuadrilla necesario para lograr la tasa de producción? ¿Cuál es la utilización resultante de la cuadrilla?

17.6-33. Un taller tiene tres máquinas idénticas que presentan fallas de cierto tipo. Se cuenta con un sistema de mantenimiento para realizar las operaciones de mantenimiento (recarga) que requiere una máquina que falla. El tiempo necesario para realizar cada operación tiene distribución exponencial con media de 30 minutos. Con probabilidad de $\frac{1}{3}$, esta operación debe realizarse una segunda vez (con la misma distribución del tiempo) para dejar la máquina en un estado operativo satisfactorio. El sistema de mantenimiento trabaja sólo en una máquina que falla a la vez; además, realiza todas las operaciones (una o dos) que requiere esa máquina sobre la base de primero en entrar, primero en salir. Después de reparada, el tiempo hasta que la máquina vuelve a fallar tiene distribución exponencial con media de 3 horas.

a) ¿Cómo deben definirse los estados del sistema para formularlo como cadena de Markov de tiempo continuo? (*Sugerencia:* Dado que se está realizando la primera operación en una máquina que falló, terminar *con éxito* esta operación y terminarla *sin éxito* son dos eventos independientes. Use la propiedad 6 para desagregar la distribución exponencial.)

b) Construya el diagrama de tasas correspondiente.

c) Desarrolle las ecuaciones de balance.

17.7-1.* Considere el modelo *M*/*G*/1.

a) Compare el tiempo esperado de espera en la cola si la distribución del tiempo de servicio es *i*) exponencial, *ii*) constante y *iii*) de Erlang con la cantidad de variación (es decir, la desviación estándar) a la mitad entre los casos constante y exponencial.

b) ¿Cuál es el efecto sobre el tiempo de espera esperado en la cola y sobre la longitud esperada de la cola si λ y μ se duplican y la escala de la distribución de tiempos de servicio se cambia de acuerdo con estos datos?

17.7-2. Considere el modelo *M*/*G*/1 con $\lambda = 0.2$ y $\mu = 0.25$.

T ***a***) Use la plantilla de Excel para este modelo (o cálculos a mano) para encontrar las medidas de desempeño L, L_q, W y W_q de los siguientes valores de σ: 4, 3, 2, 1, 0.

b) ¿Cuál es la razón de L_q con $\sigma = 4$ entre L_q con $\sigma = 0$? ¿Qué dice esto de la importancia de reducir la variabilidad de los tiempos de servicio?

c) Calcule la reducción en L_q cuando σ se reduce de 4 a 3, de 3 a 2, de 2 a 1 y de 1 a 0. ¿Cuál es la reducción más grande? ¿Cuál es la menor?

d) Use prueba y error en la plantilla para aproximar cuánto debe aumentar μ con $\sigma = 4$ para lograr la misma L_q que con $\mu = 0.25$ y $\sigma = 0$.

17.7-3. Diga si las siguientes afirmaciones sobre un sistema de colas *M*/*G*/1, donde σ^2 es la variancia de los tiempos de servicio. Establezca si es falsa o verdadera y justifique su respuesta.

a) Si se incrementa σ^2 (con λ y μ fijos) aumentarán L_q y L, pero W_q y W no cambiarán.

b) Cuando se elige entre una tortuga (μ y σ^2 pequeños) y una liebre (μ y σ^2 grandes) como servidor, la tortuga siempre gana porque proporciona una L_q pequeña.

c) Con λ y μ fijos, el valor de L_q con distribución de tiempos de servicio exponencial es el doble que cuando los tiempos de servicio son constantes.

d) Entre las distribuciones posibles (con λ y μ fijos), la distribución exponencial da el valor más grande de L_q.

17.7-4. Marsha despacha café exprés. Los clientes siguen un proceso de Poisson con tasa media de 25 por hora. El tiempo necesario para

que Marsha sirva a un cliente tiene distribución exponencial con media de 90 segundos.

a) Con el modelo *M/G/*1 encuentre L, L_q, W y W_q.

b) Suponga que sustituyen a Marsha por una máquina expendedora que requiere exactamente 90 segundos de operación por cliente. Encuentre L, L_q, W y W_q.

c) ¿Cuál es la razón de L_q en *b*) entre L_q en *a*)?

T ***d)*** Utilice por prueba y error la plantilla de Excel del modelo *M/G/*1 para aproximar cuánto debe Marsha reducir su tiempo de servicio para lograr la misma L_q que con la máquina expendedora.

17.7-5. Antonio tiene un taller de reparación de calzado. Los pares de zapatos que llegan para reparación siguen un proceso de Poisson con tasa media de 1 par por hora. El tiempo de reparación por zapato individual tiene distribución exponencial con media de 15 minutos.

a) Considere la formulación de este sistema de colas si se toma cada zapato (y no el par) como un cliente. Construya un diagrama de tasas y desarrolle las ecuaciones de balance, pero no resuelva el sistema.

b) A continuación, considere la formulación de este sistema de colas cuando los pares de zapatos se consideran los clientes. Identifique el modelo específico que se ajusta a esta formulación.

c) Calcule el número esperado de pares en el taller.

d) Calcule el tiempo esperado desde que el cliente deja un par hasta que está reparado y listo para entregarlo.

T ***e)*** Use la plantilla de Excel correspondiente para verificar sus respuestas a los incisos *c*) y *d*).

17.7-6.* Una base de mantenimiento de Friendly Skies Airline tiene instalaciones para la reparación general de un solo motor de avión a la vez. Para poner los aviones descompuestos en operación lo más pronto posible, la política ha sido alternar la reparación general de los cuatro motores de cada avión. En otras palabras, sólo se repara un motor cada vez que un avión llega a la base. Con esta política, los aviones llegan de acuerdo con un proceso de Poisson a una tasa media de 1 al día. El tiempo que se requiere para reparar un motor (una vez iniciado el servicio) tiene una distribución exponencial con media de $\frac{1}{2}$ día.

Se ha propuesto cambiar esta política por la de reparar los cuatro motores de cada avión consecutivamente cada vez que una aeronave llegue a la base. Aunque este procedimiento cuadruplicaría el tiempo esperado de servicio, cada avión iría a la base sólo la cuarta parte de las veces.

La administración debe decidir si debe continuar con este sistema o adoptar la propuesta. El objetivo es minimizar el tiempo promedio de vuelo que la flota completa pierde por día debido a las reparaciones generales de los motores.

a) Compare las dos alternativas respecto del tiempo promedio de vuelo perdido por avión cada vez que llega a la base de mantenimiento.

b) Compare las dos alternativas respecto del número promedio de aviones que pierden tiempo de vuelo por estar en la base.

c) ¿Cuál de estas comparaciones es la adecuada para que la administración tome la decisión? Explique su respuesta.

17.7-7. Reconsidere el problema 17.7-6. La administración adoptó la propuesta pero desea analizar el nuevo sistema de colas.

a) ¿Cómo debe definirse el estado del sistema para formular el modelo como cadena de Markov?

b) Construya el diagrama de tasas correspondiente.

17.7-8. En el área de producción, la compañía McAllister tiene *dos* depósitos de herramientas cada uno con *un* dependiente. Un depósito maneja sólo las herramientas para maquinaria pesada y el otro el resto. En cada uno de ellos, los mecánicos llegan a pedir herramientas con una tasa media de 18 por hora y el tiempo de servicio esperado es de 3 minutos.

Como los mecánicos se quejan de que al ir a los depósitos deben esperar demasiado, se propone combinarlos para que cualquier dependiente maneje cualquier tipo de herramienta que solicite. Se piensa que la tasa media de llegadas combinadas al depósito se duplicará a 36 por hora y que el tiempo esperado de servicio seguirá en 3 minutos. Sin embargo, no se conoce la *forma* de la distribución de probabilidad de los tiempos entre llegadas y de servicio, y no está claro qué modelo de colas es el más apropiado.

Compare el estado actual y la propuesta respecto del número esperado de mecánicos en el (los) depósito(s) de herramientas y del tiempo esperado de espera (con servicio) de cada mecánico. Con este fin, tabule los datos de los cuatro modelos de las figuras 17.6, 17.8, 17.10 y 17.11 (use $k = 2$ cuando se trate de una distribución de Erlang).

17.7-9.* Considere un sistema de colas de un servidor con entradas de Poisson, tiempos de servicio de Erlang y cola finita. En particular, suponga que $k = 2$, la tasa media de llegadas es de 2 clientes por hora, el tiempo de servicio esperado es de 0.25 horas y el máximo número de clientes que admite el sistema es 2. Este sistema se puede formular como cadena de Markov de tiempo continuo al dividir cada tiempo de servicio en dos etapas consecutivas, cada una con distribución exponencial con media de 0.125 horas, y al definir el estado del sistema como (n, p), donde n es el número de clientes en el sistema ($n = 0$, 1, 2) y p indica la etapa del servicio en la que se encuentra el cliente ($p = 0$, 1, 2, donde $p = 0$ significa que ningún cliente está en servicio).

a) Construya el diagrama de tasas correspondiente. Escriba las ecuaciones de balance y úselas para obtener la distribución de estado estable del estado de esta cadena de Markov.

b) Utilice la distribución de estado estable que obtuvo en el inciso *a*) para identificar la distribución de estado estable del número de clientes en el sistema (P_0, P_1, P_2) y el número esperado de clientes en el sistema en el estado estable (L).

c) Compare los resultados de *b*) con los que se obtendrían si la distribución de tiempos de servicio fuese exponencial.

17.7-10. Considere el modelo $E_2/M/1$ con $\lambda = 4$ y $\mu = 5$, que se puede formular como cadena de Markov de tiempo continuo al dividir cada tiempo entre llegadas en dos etapas consecutivas, con distribución exponencial con media de $1/(2\lambda) = 0.125$ y al definir el estado del sistema como (n, p), donde n es el número de clientes en el sistema ($n = 0, 1, 2, \ldots$) y p indica la etapa de la siguiente llegada (que todavía no está en el sistema) ($p = 1, 2$).

Construya el diagrama de tasas correspondiente (pero no resuelva).

17.7-11. Una compañía tiene un técnico que da servicio a un grupo grande de máquinas que están en operación. Si considera a este grupo como una fuente de entrada infinita, las descomposturas individuales son de Poisson con tasa media de 1 por hora. Para cada descompostura existe una probabilidad de 0.9 de que sólo necesite una reparación menor, en cuyo caso el tiempo para repararla tiene distribución exponencial con media de $\frac{1}{2}$ hora. De otra manera, se requiere una reparación mayor, y el tiempo para repararla tiene distribución exponencial con media de 5 horas. Como ambas distribuciones *condicio-*

nales son exponenciales, la distribución *incondicional* (combinada) de los tiempos de reparación es *hiperexponencial*.

a) Calcule la media y la desviación estándar de esta distribución hiperexponencial. [*Sugerencia:* Use las relaciones generales de teoría de probabilidad: para cualquier variable aleatoria X y cualquier par de eventos mutuamente excluyentes, E_1 y E_2, $E(X) = E(X|E_1)P(E_1) + E(X|E_2)P(E_2)$ y $\text{var}(X) = E(X^2) - E(X)^2$.] Compare esta desviación estándar con la de una distribución exponencial que tiene esta media.

b) Calcule P_0, L_q, L, W_q y W de este sistema de colas.

c) ¿Cuál es el valor condicional de W, dado que la máquina necesita reparación mayor? ¿Reparación menor? ¿Cuál es la división de L entre máquinas que requieren las dos reparaciones? (*Sugerencia:* La fórmula de Little se aplica a las categorías individuales de máquinas.)

d) ¿Cómo se definen los estados del sistema para formularlo como cadena de Markov de tiempo continuo? (*Sugerencia:* Considere qué información se necesita además del número de máquinas descompuestas para que la distribución condicional del tiempo que falta hasta el próximo evento de cada tipo sea exponencial.)

e) Construya el diagrama de tasas correspondiente.

17.7-12. Considere la variación de cola finita del modelo $M/G/1$, donde K es el número máximo de clientes que acepta el sistema. Para $n = 1, 2, \ldots$, sea X_n la variable aleatoria del número de clientes en el sistema en el tiempo t_n cuando se completa el servicio del n-ésimo cliente (no cuenta el cliente que está saliendo). Los tiempos $\{t_1, t_2, \ldots\}$ se llaman *puntos de regeneración*. $\{X_n\}$ ($n = 1, 2, \ldots$) es una cadena de Markov de tiempo discreto y se conoce como *cadena de Markov anidada*. Estas cadenas de Markov son útiles para estudiar las propiedades de los procesos estocásticos de tiempo continuo como el modelo $M/G/1$.

Considere el caso especial donde $K = 4$, el tiempo de servicio de clientes sucesivos es una constante fija, por ejemplo 10 minutos, y la tasa media de llegadas es 1 cada 50 minutos. Por lo tanto, $\{X_n\}$ es una cadena de Markov anidada con estados 0, 1, 2, 3. (Como nunca hay más de cuatro clientes en el sistema, no puede haber más de 3 en un punto de regeneración.) El sistema se observa en las salidas sucesivas, así X_n nunca puede decrecer en más de 1. Las probabilidades de transición que resultan en un incremento de X_n se obtienen directo de la distribución Poisson.

a) Encuentre la matriz de transición de un paso de la cadena de Markov anidada. (*Sugerencia:* Obtenga la probabilidad de transición del estado 3 al estado 3; use la probabilidad de 1 o más llegadas en lugar de sólo la de 1 llegada; proceda igual con las otras transiciones al estado 3.)

b) Utilice la rutina adecuada en el área de cadenas de Markov del IOR Tutorial para encontrar las probabilidades de estado estable del número de clientes en el sistema en los puntos de regeneración.

c) Calcule el número esperado de clientes en el sistema en los puntos de regeneración y compárelos con L del modelo $M/D/1$ (con $K = \infty$ sección 17.7).

17.8-1.* Southeast Airlines es una línea aérea que da servicio en Florida. Su oficina de boletos en cierto aeropuerto de Orlando tiene un agente. Hay dos colas, una para pasajeros de primera clase y otra para los de clase turista. Cuando el agente está listo para atender otro cliente, sirve al siguiente de primera clase si hay uno. Si no lo hay, sirve al siguiente de clase turista. Los tiempos de servicio tienen distribución exponencial con media de 3 minutos para ambos tipos de clientes. En las 12 horas por día que abre la oficina, los pasajeros de primera clase llegan de manera aleatoria a una tasa media de 2 por hora y 10 por hora los de clase turista.

a) ¿Qué tipo de modelo se ajusta a este sistema de colas?

T ***b***) Encuentre las medidas medias de desempeño L, L_q, W y W_q de ambos tipos de pasajeros.

c) ¿Cuál es el tiempo de espera esperado antes de iniciar el servicio a los clientes de primera como fracción del tiempo de los de clase turista?

d) Determine el número promedio de horas por día que está ocupado el agente.

T **17.8-2.** Considere el modelo con prioridades sin interrupción de la sección 17.8. Suponga que sólo hay dos clases de prioridad, con $\lambda_1 = 2$ y $\lambda_2 = 3$. Cuando diseñe este sistema puede elegir entre dos posibilidades: 1) un servidor rápido ($\mu = 6$) y 2) dos servidores lentos ($\mu = 3$).

Compare estas alternativas con las cuatro medidas de desempeño usuales (W, L, W_q, L_q) de las clases individuales de prioridad (W_1, W_2, L_1, L_2, etc.). ¿Qué alternativa se debe preferir si la preocupación primordial es el tiempo de espera esperado en el *sistema* por la clase de prioridad 1 (W_1)? ¿Cuál es preferible si la preocupación es el tiempo esperado en la *cola* por la clase de prioridad 1?

17.8-3. Considere el modelo con prioridades sin interrupción de la sección 17.8. Suponga que hay tres clases de prioridades, con $\lambda_1 = 1$, $\lambda_2 = 1$ y $\lambda_3 = 1$. Los respectivos tiempos de servicio esperados de las clases 1, 2 y 3 son 0.4, 0.3 y 0.2; en consecuencia, $\mu_1 = 2.5$, $\mu_2 = 3\frac{1}{3}$ y $\mu_3 = 5$.

a) Calcule W_1, W_2 y W_3.

b) Repita el inciso *a*) a la aproximación de aplicar el modelo general de prioridades sin interrupción de la sección 17.8. Como el modelo supone tiempos de servicio esperados iguales para todas las clases, use un tiempo de 0.3 en lugar de $\mu = 3\frac{1}{3}$. Compare los resultados con los que obtuvo en *a*) y evalúe si la aproximación que proporciona este supuesto es buena.

T **17.8-4.*** Cierto centro de trabajo dentro de un taller se puede representar como un sistema de colas con un servidor, en el que los trabajos tienen llegadas de Poisson con tasa media de 8 por día. Aunque las llegadas son de tres tipos, el tiempo que se requiere para realizar cualquier trabajo tiene la misma distribución exponencial, con media de 0.1 días. La práctica ha sido realizarlos en el orden en que llegan, pero es importante que los trabajos del tipo 1 no esperen mucho, mientras que este detalle no importa tanto con trabajos tipo 2 y casi no tiene importancia si se trata de los del tipo 3. Los tres tipos llegan con tasas medias respectivas de 2, 4 y 2 por día. Como los tres tipos han experimentado retrasos bastante grandes en promedio, se propone seleccionarlos según una disciplina de prioridades apropiada.

Compare el tiempo de espera esperado (con servicio) de cada tipo de trabajo si la disciplina de la cola es *a*) primero en entrar, primero en salir, *b*) prioridades sin interrupción y *c*) prioridades con interrupción.

T **17.8-5.** Reconsidere el problema de la sala de urgencias del *Hospital General* de la sección 17.8. Suponga que se ha precisado la definición de las categorías de pacientes a fin de pasar los casos marginales a una categoría más baja. En consecuencia, sólo 5% de los pacientes se clasificará como caso crítico, 20% como caso serio y 75% como

caso estable. Desarrolle una tabla que muestre los datos que se presentaron en la tabla 17.4 de este problema revisado.

17.8-6. Considere el sistema del problema 17.4-6. Suponga que los clientes tipo 1 son más importantes que los tipo 2. Si se cambiara la disciplina de la cola de primero en entrar, primero en salir a un sistema de prioridades con prioridad sin interrupción de los clientes tipo 1 sobre los clientes tipo 2, ¿aumentaría, disminuiría o no cambiaría el número total de clientes en el sistema?

***a*)** Determine la respuesta sin realizar cálculos y después presente el razonamiento que lo llevó a su conclusión.

T ***b*)** Verifique su conclusión en *a*) al encontrar el número total esperado de clientes en el sistema con cada disciplina de la cola.

17.8-7. Considere el modelo de colas con disciplina de prioridades con interrupción de la sección 17.8. Suponga que $s = 1$, $N = 2$ y $(\lambda_1 + \lambda_2) < \mu$; y sea P_{ij} la probabilidad de estado estable de que haya i miembros de prioridad alta y j miembros de prioridad baja en el sistema ($i = 0, 1, 2, \ldots$; $j = 0, 1, 2, \ldots$). Use un método análogo al descrito en la sección 17.5 para derivar un sistema de ecuaciones lineales cuya solución simultánea sea P_{ij}. No obtenga esta solución.

17.9-1. Lea el artículo de referencia que describe el estudio de investigación de operaciones que se resume en la Viñeta de aplicación que se presentó en la sección 17.9. Describa de manera breve la forma como se aplicó la teoría de colas a este estudio. Después, elabore una lista de los beneficios financieros y no financieros que arrojó dicho estudio.

17.9-2. Considere un sistema de colas con dos servidores al que los clientes llegan de dos fuentes diferentes. De la fuente 1, siempre llegan 2 a la vez y el tiempo entre llegadas consecutivas de pares de clientes es exponencial con media de 20 minutos. La fuente 2 es en sí misma un sistema de colas con dos servidores que tiene entradas de Poisson con media de 7 clientes por hora y el tiempo de servicio por servidor es exponencial con media de 15 minutos. Cuando termina el servicio de un cliente en la fuente 2, de inmediato entra al sistema de colas estudiado para recibir otro servicio. En este último sistema, la disciplina de la cola es prioridades con interrupción y los clientes de la fuente 1 tienen prioridad sobre los de la fuente 2. Sin embargo, los tiempos de servicio son independientes e idénticamente distribuidos para ambos tipos de clientes según una distribución exponencial con media de 6 minutos.

***a*)** Primero estudie el problema de derivar la distribución de estado estable *sólo* del número de clientes de la fuente 1 que hay en el sistema. Use una formulación de cadenas de Markov de tiempo continuo para definir los estados y construir el diagrama de tasas que facilite el desarrollo de esta distribución (pero no la obtenga).

***b*)** A continuación, enfóquese en el problema de derivar la distribución de estado estable del número *total* de clientes de los dos tipos en el sistema de colas. Utilice una formulación de cadenas de Markov de tiempo continuo; defina los estados y construya el diagrama de tasas que facilite la derivación de esta distribución (pero no la obtenga).

***c*)** Ahora examine el problema de derivar la distribución *conjunta* de estado estable del número de clientes de cada tipo en el sistema. Use una formulación de cadenas de Markov de tiempo continuo para definir los estados y construir el diagrama de tasas para derivar esta distribución (pero no la obtenga).

17.9-3. Considere un sistema de dos colas infinitas en serie, donde cada instalación de servicio tiene un servidor. Todos los tiempos de servicio son independientes y tienen distribución exponencial con media de 3 minutos en la instalación 1 y 4 min en la instalación 2. La instalación 1 tiene entradas Poisson con tasa media de 10 por hora.

***a*)** Encuentre la distribución de estado estable del número de clientes en la instalación 1 y después en la 2. Muestre la forma de solución de producto de la distribución *conjunta* del número en las respectivas instalaciones.

***b*)** ¿Cuál es la probabilidad de que ambos servidores estén desocupados?

***c*)** Encuentre el número *total* esperado de clientes en el sistema y el tiempo de espera esperado *total* (con servicio) de un cliente.

17.9-4. Bajo los supuestos de la sección 17.9 para un sistema de colas infinitas en serie, este tipo de redes de colas es un caso especial de una red de Jackson. Demuestre que esto es cierto con la descripción del sistema como una red de Jackson, incluya la especificación de los valores de a_j y de la p_{ij}, dada λ de este sistema.

17.9-5. Considere una red de Jackson con tres instalaciones de servicio cuyos parámetros tienen los valores que se muestran en la siguiente tabla.

Instalación j	s_j	μ_j	a_j	p_{ij} $i = 1$	$i = 2$	$i = 3$
$j = 1$	1	25	6	0	0.2	0.4
$j = 2$	1	30	8	0.5	0	0.3
$j = 3$	1	20	4	0.4	0.3	0

T ***a*)** Encuentre la tasa de llegadas total de cada instalación.

***b*)** Encuentre la distribución de estado estable del número de clientes en la instalación 1, en la 2 y en la 3. Después, muestre la forma de solución de producto de la distribución conjunta del número en cada instalación.

***c*)** ¿Cuál es la probabilidad de que todas las instalaciones tengan colas vacías (sin clientes que esperan servicio)?

***d*)** Encuentre el número total esperado de clientes en el sistema.

***e*)** Encuentre el tiempo de espera esperado total (que incluye los tiempos de servicio) de un cliente.

T **17.10-1.** Al describir el análisis económico del número de servidores que debe proporcionarse en un sistema de colas, en la sección 17.10 se indica que la función objetivo es minimizar $E(\text{CT}) = C_s s + C_w L$. El propósito de este problema es explorar el efecto de los tamaños relativos de C_s y C_w sobre el número óptimo de servidores.

Suponga que el sistema de colas se ajusta al modelo *M/M/s* con $\lambda = 8$ y $\mu = 10$ clientes por hora. Use la plantilla de Excel del OR Courseware para el análisis económico con *M/M/s* y encuentre el número óptimo de servidores en cada caso.

***a*)** $C_s = \$100$ y $C_w = \$10$.

***b*)** $C_s = \$100$ y $C_w = \$100$.

***c*)** $C_s = \$10$ y $C_w = \$100$.

T **17.10-2.*** Jim McDonald, gerente del restaurante de hamburguesas McBurger, sabe que proporcionar un servicio rápido es la clave del éxito. Es posible que los clientes que esperan mucho vayan a otro lugar la próxima vez. Estima que cada minuto que un cliente tiene que esperar antes de terminar su servicio le cuesta un promedio

de 30 centavos en negocio futuro perdido. Por lo tanto, desea estar seguro de siempre tener suficientes cajas abiertas para que la espera sea mínima. Un empleado de tiempo parcial opera cada caja, obtiene la orden del cliente y cobra. El costo total de cada empleado es de $9 por hora.

Durante la hora del almuerzo, los clientes llegan según un proceso de Poisson a una tasa media de 66 por hora. Se estima que el tiempo necesario para servir a un cliente tiene distribución exponencial con media de 2 minutos.

Determine cuántas cajas debe abrir Jim durante este tiempo para minimizar su costo total esperado por hora.

T **17.10-3.** La compañía Garret-Tompkins tiene tres copiadoras para uso de sus empleados. Sin embargo, debido a quejas recientes de la cantidad de tiempo que pierden en espera de que se desocupe una copiadora, la gerencia planea agregar una o más.

Durante las 2 000 horas de trabajo al año, los empleados llegan al área de copiado según un proceso de Poisson con tasa media de 40 por hora. Se cree que el tiempo que cada empleado necesita una copiadora tiene distribución exponencial con media de 4 minutos. El costo promedio de la productividad perdida debida al tiempo que pasa un empleado en el área de copiado se estima en $40 por hora. La renta de cada copiadora es de $4 000 por año.

Determine cuántas copiadoras debe tener la compañía para minimizar su costo total esperado por hora.

17.11-1. De la parte inferior de las referencias seleccionadas que se proporcionan al final del capítulo, elija una de estas aplicaciones de teoría de colas ganadoras de premios. Lea este artículo y después escriba un resumen de dos páginas de la aplicación y los beneficios (financieros y no financieros) que ésta proporciona.

17.11-2. De la parte inferior de las referencias seleccionadas que se proporcionan al final del capítulo, elija tres de ellas que se refieran a aplicaciones de teoría de colas ganadoras de premios. En el caso de cada una de ellas, lea el artículo y después escriba un resumen de una página de la aplicación y los beneficios (financieros y no financieros) que ésta proporciona.

CASOS

CASO 17.1 Reducción de inventario en proceso

Jim Wells, vicepresidente de manufactura en Northern Airplane Company, está alterado. Su caminata de esta mañana por la planta más importante de la compañía lo dejó de mal humor. Sin embargo, ahora puede descargar su temperamento en Jerry Carstairs, el gerente de producción de la planta, a quien acaba de llamar a su oficina.

"Jerry, acabo de regresar de recorrer la planta y estoy muy enojado." "¿Cuál es el problema, Jim?" "Bueno, sabes cuánto he recalcado la necesidad de disminuir nuestro inventario en proceso." "Sí, hemos trabajado duro en eso", responde Jerry. "¡Pues no lo suficiente!" Jim sube más la voz. "¿Sabes lo que encontré junto al proceso?" "No." "Cinco hojas de metal esperando que las formen en secciones de alas. Después, exactamente en el área siguiente de inspección, ¡13 secciones de alas! El inspector estaba revisando una de ellas, pero las otras 12 sólo estaban ahí. Tú sabes que tenemos un par de cientos de miles de dólares atados en cada una de esas secciones de ala, es decir, tenemos algunos millones de dólares en metal terriblemente costoso sólo en espera. ¡No podemos tener eso!"

El consternado Jerry Carstairs intenta responder. "Sí, Jim, estoy consciente de que la estación de inspección es un cuello de botella. Casi nunca está tan mal como la encontraste esta mañana." "Eso espero, claro", replica Jim, "pero necesitas prevenir que algo tan perjudicial ocurra ni siquiera en ocasiones. ¿Qué propones hacer al respecto?" Es notorio que ahora Jerry se anima al responder. "Bueno, en realidad, ya he estado trabajando en este problema; tengo un par de propuestas en la mesa y ya pedí a un analista de IO, que es parte de mi personal, que estudie estas propuestas y regrese con recomendaciones." "Bien", dice Jim, "me da gusto ver que tratas de resolverlo. Dale a esto tu prioridad más alta y repórtame cuanto antes". "Está bien", promete Jerry.

El siguiente es el problema que Jerry y su analista de IO estudian. Cada una de 10 prensas idénticas se usa para formar las secciones de alas a partir de hojas de metal con un procesado especial. Las hojas llegan de manera aleatoria al grupo de prensas con una tasa media de 7 por hora. El tiempo que requiere la prensa para formar una sección de ala a partir de la hoja de metal tiene distribución exponencial con media de 1 hora. Cuando termina, la sección de ala llega de manera aleatoria a una estación de inspección con la misma tasa media que las hojas llegan a las prensas (7 por hora). Un solo inspector tiene el trabajo de tiempo completo de inspeccionar las secciones de ala para asegurar que cumplen con las especificaciones. Cada inspección toma 7 y1/2 minutos, de manera que puede inspeccionar 8 secciones por hora. Esta tasa de inspección ha dado como resultado una cantidad promedio sustancial de inventario en proceso en la estación de inspección (esto es, el número promedio de secciones de ala que esperan pasar la inspección es bastante grande), además del que se encuentra en el grupo de máquinas.

El costo de este inventario en proceso se estima en $8 por hora por cada hoja de metal en las prensas o cada sección de ala en la estación de inspección. Por lo tanto, Jerry Castairs tiene dos propuestas alternativas para reducir el nivel de inventario en proceso.

La propuesta 1 es usar un poco menos de potencia en las prensas (lo que aumentaría su tiempo promedio para formar una sección de ala a 1.2 horas), de manera que el inspector pudiera mantener el paso con su producción. Esta medida también reduciría el costo de energía para operar las máquinas de $7.00 a $6.50 por hora. (Por el contrario, incrementar la potencia al máximo aumentaría este costo a $7.50 por hora y disminuiría el tiempo promedio para formar cada sección de ala a 0.8 horas.)

La propuesta 2 es sustituir el inspector por uno más joven para esta tarea. Éste es un poco más rápido (aunque con alguna

variabilidad en sus tiempos de inspección porque tiene menos experiencia), por lo que tendría un mejor paso. (Su tiempo de inspección tendrá una distribución de Erlang con media de 7.2 minutos y parámetro de forma $k = 2$.) Este inspector se encuentra en una clasificación de puestos que requiere una compensación total (incluyendo prestaciones) de $19 por hora, mientras que el inspector actual tiene una clasificación más baja y gana $17 por hora. (Los tiempos de inspección de cada inspector son los normales en esa categoría en la clasificación.)

Usted es el analista de IO en el equipo de Jerry Castairs a quien le han pedido que estudie este problema. Él quiere “que utilice las últimas técnicas en IO para ver cuánto disminuiría cada propuesta el inventario en proceso para hacer sus recomendaciones.”

***a*)** Para proporcionar una base de comparación, comience por evaluar el estado actual. Determine la cantidad esperada de inventario en proceso en las prensas y en la estación de inspección. Después, calcule el costo total esperado por hora, pero tome en cuenta lo siguiente: el costo del inventario en proceso, el costo de la energía para operar las prensas y el costo del inspector.

***b*)** ¿Cuál sería el efecto de la propuesta 1? ¿Por qué? Realice comparaciones específicas con los resultados del inciso *a*). Explique esta comparación a Jerry Castairs.

***c*)** Determine el efecto de la propuesta 2. Realice comparaciones específicas con los resultados del inciso *a*). Explique esta comparación a Jerry Castairs.

***d*)** Presente sus recomendaciones para reducir el nivel de inventario en proceso en la estación de inspección y el grupo de máquinas. Sea específico en sus recomendaciones y apóyelas con un análisis cuantitativo como el que realizó en el inciso *a*). Realice comparaciones específicas con los resultados del inciso *a*) y cite las mejoras que serían el resultado de sus sugerencias.

RESUMEN DE LOS CASOS ADICIONALES EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

CASO 17.2 Dilema de colas

Muchos clientes molestos se han quejado por las largas esperas que deben hacer para acceder a un centro de llamadas. Parece que son necesarios más representantes de servicio para contestar las llamadas. Otra opción es entrenar a los representantes para que contesten las llamadas de manera más eficiente. Se han propuesto algunos criterios para definir los niveles satisfactorios de servicio. Es necesario aplicar la teoría de colas para determinar cómo debe rediseñarse la operación del centro de llamadas.

CAPÍTULO 18

Teoría de inventarios

"Lo siento, no tenemos ese artículo." ¿Cuántas veces ha escuchado esta respuesta cuando va de compras? En muchos casos, se encuentran negocios que no hacen un buen trabajo al administrar sus *inventarios* (bienes almacenados para uso o venta futuros). No colocan sus pedidos de reabastecimiento con suficiente anticipación para evitar faltantes. Estos negocios se pueden beneficiar de los tipos de técnicas de la administración científica de inventarios que se describen en este capítulo.

No sólo los comerciantes deben administrar inventarios. En realidad, los inventarios prevalecen en el mundo de los negocios. Mantenerlos en un buen nivel es necesario para las compañías que operan con productos físicos, como fabricantes, distribuidores y comerciantes. Por ejemplo, los fabricantes necesitan contar con inventarios de materiales que se requieren para la manufactura de productos. También deben almacenar productos terminados en espera de ser enviados. De manera similar, tanto los distribuidores como las tiendas deben mantener inventarios de bienes disponibles para cuando los consumidores los soliciten.

El costo asociado con almacenar ("mantener") inventarios es también muy alto, quizá un cuarto del valor del inventario. Por lo tanto, los costos en los que se incurre al guardar inventarios en Estados Unidos ascienden a cientos de miles de millones de dólares anuales. Reducir los costos de almacenamiento para evitar inventarios innecesariamente grandes puede mejorar la competitividad de cualquier empresa.

Algunas compañías japonesas han sido pioneras en la introducción de los *sistemas de inventarios justo a tiempo* (un sistema que hace hincapié en la planeación y programación para que los materiales necesarios lleguen "justo a tiempo" para su uso). Se han logrado grandes ahorros mediante la reducción de los niveles de inventarios a un mínimo.

Muchas compañías de otros países también han renovado la manera en que manejan sus inventarios. La aplicación de las técnicas de IO en esta área (en ocasiones llamadas *administración científica de inventarios*) proporciona una herramienta poderosa para lograr una ventaja competitiva.

¿Cómo utilizan las compañías la investigación de operaciones para mejorar sus **políticas de inventarios** respecto a cuándo y cuánto reabastecer su inventario? La **administración científica de inventarios** comprende los siguientes pasos:

1. Formular un *modelo matemático* que describa el comportamiento del sistema de inventarios.
2. Elaborar una política *óptima* de inventarios a partir de ese modelo.
3. Utilizar un *sistema de procesamiento de información* computarizado para mantener registros de los niveles del inventario.
4. A partir de estos registros, utilizar la política óptima de inventarios para señalar cuándo y cuánto conviene reabastecer.

Los modelos matemáticos de inventarios que aplican este enfoque se pueden dividir en dos grandes categorías: modelos determinísticos y modelos estocásticos, según la posibilidad de *predecir la demanda*. La **demanda** de un producto en inventario es el número de unidades que será necesario extraer de éste para algún uso (como venta) durante un periodo específico. Si la demanda

en periodos futuros se puede pronosticar con precisión considerable, es razonable usar una política de inventarios que suponga que los pronósticos siempre serán muy precisos. Éste es el caso de la *demanda conocida*, ante la cual se usa un modelo de inventarios *determinístico*. Sin embargo, cuando no se puede predecir con exactitud, es necesario usar un modelo de inventarios *estocástico*, en el cual la demanda en cualquier periodo es una variable aleatoria en lugar de una constante conocida.

Existen varias consideraciones básicas relacionadas con la determinación de una política de inventarios que deben reflejarse en el modelo matemático, las cuales se ilustran en los ejemplos que se presentan en la primera sección y después se describen en términos generales en la sección 18.2. En la sección 18.3 se desarrollan y analizan modelos de inventarios determinísticos para enfrentar situaciones en las que el nivel de inventario está sujeto a una revisión continua. En la sección 18.4 se utiliza el mismo enfoque para manejar situaciones donde la planeación se realiza para cubrir una serie de periodos y no de manera continua. En la sección 18.5 se despliegan ciertos modelos determinísticos para coordinar los inventarios en varios puntos a lo largo de la cadena de suministros de una compañía. Las siguientes dos secciones presentan modelos estocásticos, primero de revisión continua y después para un solo periodo. (En el complemento de este capítulo que se encuentra en el sitio en internet de este libro se presenta el modelo estocástico de revisión periódica para periodos múltiples.) La sección 18.8 presenta un área relativamente nueva de la teoría de inventarios llamada *administración de la ganancia* que se relaciona con la maximización de las ganancias esperadas de una compañía cuando ésta comercializa un tipo especial de productos perecederos cuyo inventario total se le debe brindar al usuario en un determinado momento o, de otra forma, se pierde para siempre. (Ciertas industrias de servicios como, por ejemplo, una compañía aérea que ofrece su inventario total de asientos en un determinado vuelo en el momento específico del vuelo, hace un uso extensivo de la administración de las ganancias.)

18.1 EJEMPLOS

Se presentarán dos ejemplos en contextos bastante diferentes (un fabricante y un distribuidor) para los que es necesario desarrollar una política de inventarios.

EJEMPLO 1 Fabricación de bocinas para televisores

Una compañía que fabrica televisores produce sus propias bocinas para utilizarlas en la fabricación de estos aparatos. Los televisores se ensamblan en una línea de producción continua a una tasa de 8 000 por mes, y se necesita una bocina por televisor. Las bocinas se fabrican por lotes, pues no justifican toda una línea de producción y se pueden producir cantidades relativamente grandes en un tiempo corto. Por lo tanto, estas bocinas se colocan en inventario hasta que se necesitan para ser ensambladas en los televisores. La compañía está interesada en determinar cuándo producir un lote de bocinas y cuántas producir en cada lote. Es necesario tomar en cuenta varios costos:

1. Cada vez que se produce un lote, se incurre en un **costo de preparación** de 12 000 dólares. Esta cantidad incluye el "costo de preparar las máquinas y herramientas", los costos administrativos, los de registros, etc. Observe que la existencia de estos costos es un argumento para producir lotes grandes de bocinas.
2. El **costo unitario de producción** de una sola bocina (excluye el costo de preparación) es de 10 dólares independientemente del tamaño del lote fabricado. (No obstante, en general, el costo unitario de producción no necesita ser constante y puede decrecer junto con el tamaño del lote.)
3. La producción de bocinas en grandes lotes conduce a la formación de un inventario grande. La estimación del **costo de mantener** una bocina en almacén es de $0.30 por mes. Este monto incluye el costo de capital comprometido en el inventario. Como el dinero invertido en él no se puede usar de otra manera productiva, este costo de capital consiste en el rendimiento perdido (llamado *costo de oportunidad*) porque debe prescindirse de usarlo de otra forma. Otros componentes del costo de mantener inventarios incluyen el costo de renta del espacio de almacén, los seguros de incendio, robo o vandalismo, impuestos basados en el valor del inventario y el costo de personal que supervisa y protege el inventario.

4. La política de la compañía prohíbe la planeación deliberada de faltantes de cualquiera de sus componentes. Sin embargo, en ocasiones faltan bocinas y se estima que cada una de ellas que falta cuando se necesita cuesta $1.10 por mes. Este **costo por faltantes** incluye el costo de instalar las bocinas con el televisor totalmente ensamblado, el interés perdido por el retraso en recibir ingresos por ventas, el costo de mantener registros y otros.

Se desarrollará la política de inventarios para este ejemplo con la ayuda del primer modelo de inventario presentado en la sección 18.3.

EJEMPLO 2 Distribuidor mayorista de bicicletas

Un distribuidor al mayoreo de bicicletas tiene problemas con faltantes de un modelo popular (una bicicleta pequeña, de una velocidad, para niñas) por lo que en la actualidad revisa su política de inventarios con respecto a este modelo. El distribuidor compra este modelo al fabricante cada mes y después surte las bicicletas a distintas tiendas del oeste de Estados Unidos para satisfacer sus órdenes de compra. Existe incertidumbre sobre cuál será la demanda de bicicletas por parte de las tiendas en cualquier mes. En consecuencia, la pregunta es: ¿cuántas bicicletas debe ordenar al fabricante en un mes determinado, dado el nivel de inventario al comenzar ese mes?

El distribuidor analizó sus costos y determinó que los factores importantes son:

1. El *costo por ordenar*, es decir, el costo de colocar un pedido más el costo de las bicicletas que compra, tiene dos componentes: el costo del trabajo administrativo necesario para colocar la orden se estima en 2 000 dólares, mientras que, para él, el costo real de cada bicicleta es de 350 dólares.
2. El *costo de mantener*, es decir, el costo de tener un inventario, es de 10 dólares por cada bicicleta que queda al final del mes. Este monto representa el capital comprometido, espacio de almacén, seguros, impuestos, etcétera.
3. El *costo por faltantes* es el costo por no tener una bicicleta disponible cuando se necesita. La mayoría de los modelos se puede reordenar al fabricante sin problemas y por lo general los negocios aceptan surtidos atrasados. Aun así, el distribuidor siente que incurre en una pérdida, que ha estimado en 150 dólares por mes por bicicleta faltante. Este costo toma en cuenta la pérdida posible de ventas futuras por el deterioro de la imagen. Otros componentes de este costo incluyen el interés perdido por el retraso de los ingresos por ventas y los costos administrativos adicionales asociados con los faltantes. Si algunas tiendas cancelaran sus órdenes por los retrasos, los ingresos por ventas perdidas deberán incluirse en el costo de faltantes. Por fortuna, estas cancelaciones no ocurren en el caso de este distribuidor.

Se regresará a este ejemplo en la sección 18.7.

Estos ejemplos muestran las dos posibilidades que existen de que la empresa *reabastezca su inventario*, lo cual depende de la situación. Una posibilidad es que la empresa *produzca* las unidades necesarias (como el fabricante de televisores que produce bocinas). La otra es que *ordene* las unidades a un proveedor (como el distribuidor de bicicletas que hace el pedido al fabricante). Los modelos de inventarios no necesitan distinguir entre estas dos formas de reabastecimiento del inventario, por lo que se usarán los términos *producir* u *ordenar* como sinónimos.

Ambos ejemplos manejan un producto específico (bocinas para cierto tipo de televisores o un modelo especial de bicicleta). En la mayor parte de los modelos de inventarios se considera sólo un producto a la vez. Todos los modelos de inventarios que se presentan aquí suponen un solo producto.

Ambos ejemplos indican que existe una compensación entre los costos involucrados. La siguiente sección estudia las componentes de costo básicas de los modelos de inventarios para determinar la compensación óptima entre estos costos.

18.2 COMPONENTES DE LOS MODELOS DE INVENTARIOS

Debido a que las políticas de inventarios afectan las ganancias, la elección entre una política y otra depende de su rentabilidad relativa. Como ya se vio en los ejemplos 1 y 2, algunos de los costos que determinan esta rentabilidad son 1) los costos de ordenar o fabricar, 2) los costos de mantener o almacenar y 3) los costos de penalización por faltantes o demanda insatisfecha. Otros costos relevantes incluyen 4) los ingresos, 5) los costos de recuperación y 6) las tasas de descuento. A continuación se describirán estos seis factores.

El **costo de ordenar** o fabricar una cantidad z (ya sea mediante *compra* o *producción de esa cantidad*) se puede representar por una función $c(z)$. La forma más sencilla de esta función es aquella que es directamente proporcional a la cantidad ordenada o producida, es decir, $c \cdot z$, donde c representa el precio unitario pagado. Otro supuesto común es que $c(z)$ se compone de dos partes: un término que es directamente proporcional a la cantidad ordenada o producida y un término que es una constante K para z positiva y 0 para $z = 0$. En este caso,

$$c(z) = \text{costo de ordenar } z \text{ unidades}$$
$$= \begin{cases} 0 & \text{si } z = 0 \\ K + cz & \text{si } z > 0, \end{cases}$$

donde K = costo fijo o de preparación y c = costo unitario.

La constante K incluye el costo administrativo de ordenar o, cuando se fabrica, el costo del trabajo de preparación para poner en marcha la producción.

Existen otros supuestos que se pueden hacer respecto del costo de ordenar o fabricar, pero este capítulo se limita a los casos que se acaban de describir.

En el ejemplo 1 se fabrican bocinas y el costo fijo o de preparación de la corrida de producción es de 12 000 dólares. Además, cada bocina cuesta 10 dólares; el costo de *producción* cuando se ordena una corrida de producción de z bocinas está dado por

$$c(z) = 12\,000 + 10z, \qquad \text{para } z > 0.$$

En el ejemplo 2, el distribuidor hace un pedido de bicicletas al fabricante, y el costo de *ordenar* está dado por

$$c(z) = 2\,000 + 350z, \qquad \text{para } z > 0.$$

El **costo de mantener inventario** (a veces llamado *costo de almacenar*) representa los costos asociados con el almacenamiento del inventario hasta que se vende o se usa. Este costo incluye el costo del capital invertido, espacio, seguros, protección e impuestos atribuibles al almacenamiento. Desde otra perspectiva, se puede evaluar de manera continua o por periodo. En este caso puede ser una función de la cantidad máxima que se guarda durante un periodo, de la cantidad promedio en el almacén o de la cantidad en inventario al final del periodo. Este último punto de vista se adoptará a lo largo de este capítulo.

En el ejemplo de las bicicletas, el costo de mantener es de 10 dólares por bicicleta que queda al final del mes. En el ejemplo de las bocinas de TV, dicho costo se calcula en forma continua como \$0.30 por bocina en inventario por mes, o sea, el costo de mantener promedio por mes es de \$0.30 por el número promedio de bocinas en el inventario mensual.

El **costo por faltantes** (a veces llamado *costo de demanda insatisfecha*) surge cuando la cantidad que se requiere de un bien (demanda) es mayor que el inventario disponible. Este costo depende de cuál de los dos casos siguientes se aplica.

En un caso, llamado **con faltantes**, la demanda excesiva no se pierde, sino que queda pendiente hasta que se pueda satisfacer con el siguiente reabastecimiento normal. Para una empresa que incurre en un faltante temporal para cumplir con sus clientes (como en el ejemplo de las bicicletas), el costo por faltantes se puede interpretar como la pérdida de la imagen ante los clientes debido al retraso, su duda para realizar negocios subsecuentes con la empresa, el costo del ingreso retrasado y el trabajo administrativo adicional. En el caso de un fabricante que incurre en un faltante temporal de materiales necesarios para la producción (como un faltante de bocinas para el ensamble de los

televisores), el costo por faltantes se convierte en el costo asociado al retraso en la terminación del proceso de producción.

En el segundo caso, llamado **sin faltantes**, si ocurre un exceso de demanda sobre el inventario disponible, el distribuidor no puede esperar a la siguiente entrega normal para reabastecer el inventario, ya sea que 1) el exceso de demanda se satisfaga mediante un envío prioritario o 2) no se cumpla todo porque las órdenes fueron canceladas. En la situación 1, el costo por faltantes se puede interpretar como el costo del envío prioritario. En la situación 2, este costo por faltantes se puede ver como la pérdida en la que se incurre por no satisfacer la demanda, más el costo de perder negocios futuros debido a la pérdida de la imagen.[1]

El **ingreso** puede o no incluirse en el modelo. Si se supone que el mercado establece tanto el precio como la demanda de un producto y por ello ambos factores están fuera del control de la compañía, el rendimiento sobre las ventas (si se cumple la demanda) es independiente de la política de inventarios de la compañía y puede dejarse fuera; pero si no se incluye en el modelo, entonces la *pérdida del ingreso* debe incluirse en el costo de penalización por faltantes siempre que la empresa no pueda cumplir con esa demanda y se pierda la venta. Lo que es más, aun en el caso de que se permitan faltantes, debe incluirse el costo del retraso en el ingreso dentro del costo por faltantes. Con estas interpretaciones, en el resto del capítulo el ingreso o rendimiento no se considerará en forma explícita.

El **valor de rescate** o salvamento de un producto es el valor de un artículo sobrante cuando no se requiere más del inventario. Para la empresa, el valor de rescate representa el valor de desecho del artículo, quizá a través de una venta con descuento. El negativo del valor de rescate se llama **costo de recuperación**. Si existe un costo asociado al hecho de poder deshacerse de un artículo, el costo de recuperación puede ser positivo. Se supondrá en adelante que cualquier costo de recuperación se incorpora al *costo de mantener*.

Por último, la **tasa de descuento** toma en cuenta el valor del dinero en el tiempo. Cuando una empresa compromete capital en inventarios, no puede usarlo para otros fines. Por ejemplo, podría invertirlo en algo seguro, como bonos gubernamentales, y obtener un rendimiento sobre la inversión dentro de un año, por ejemplo de 7%. Entonces, 1 dólar invertido hoy valdría \$1.07 en un año; dicho de otra manera, la ganancia anual dentro de un año de 1 dólar es equivalente a $\alpha = \$1/\$1.07$ hoy, cantidad que se conoce como factor de descuento. Así, al considerar la productividad de una política de inventarios, la ganancia o los costos calculados para dentro de un año deben multiplicarse por α, para dentro de dos años, por α^2; etc. (También pueden usarse unidades de tiempo diferentes a un año.) El beneficio total calculado de esta forma se conoce como *valor presente neto*.

En problemas que tienen un horizonte de planeación corto, puede suponerse que α es igual a 1 (y puede despreciarse) puesto que el valor corriente de 1 dólar no cambia mucho en este corto tiempo. Sin embargo, en los problemas con horizontes lejanos debe incluirse el factor de descuento.

Al usar técnicas cuantitativas para buscar políticas óptimas de inventarios, se utiliza el criterio para minimizar el costo descontado total (esperado). Bajo los supuestos de que el precio y la demanda del producto no se encuentran bajo el control de la compañía y que el ingreso perdido o retrasado se incluye entre los costos de penalización por faltantes, minimizar el costo equivale a maximizar el ingreso neto. Otro criterio útil que se debe tomar en cuenta es que la política de inventarios que se determine debe ser sencilla, es decir, la regla que indica *cuándo* y *cuánto conviene ordenar* debe ser de fácil comprensión y sencilla de implantar. La mayor parte de las políticas que se estudian en este capítulo poseen esta propiedad.

Como se mencionó al principio del capítulo, los modelos de inventarios se clasifican en *determinísticos* o *estocásticos* según si se conoce la demanda del periodo o si se trata de una variable aleatoria que tiene una distribución de probabilidad conocida. La producción de bocinas por lotes del ejemplo 1 de la sección 18.1 es un ejemplo de demanda determinística pues se supone que se usan en el ensamble de los televisores a una tasa fija de 8 000 al mes. La compra de bicicletas al distribuidor por parte de las tiendas en el ejemplo 2 de la sección 18.1 ilustra una demanda aleatoria

[1] Un análisis de la situación 2 se describe en E. T. Anderson, G. J. Fitzsimons y D. Simester, "Measuring and Mitigating the Costs of Stockouts", en *Management Science*, **52**(11): 1751-1763, noviembre de 2006. Un análisis acerca de si los inventarios con faltantes o sin faltantes proporcionan una opción menos costosa en determinadas circunstancias se puede consultar en B. Janakiraman, S. Seshadri y J. G. Shanthikumar, "A Comparison of the Optimal Costs of Two Canonical Inventory Systems", en *Operations Research*, **55**(5): 866-875, septiembre-octubre de 2007.

porque la demanda total mensual varía de un mes a otro según alguna distribución de probabilidad. Otra componente del modelo de inventarios es el **tiempo de entrega**, que es el lapso que transcurre desde que se coloca una orden de reabastecimiento (ya sea por compra o producción) hasta la recepción de los bienes. Si el tiempo de entrega es siempre el mismo (*fijo*), el reabastecimiento se puede programar justo cuando se desea. La mayoría de los modelos del capítulo suponen que esto ocurre, ya sea porque la entrega es casi instantánea o porque se conoce cuándo se necesitará y el tiempo de entrega es fijo.

Otra clasificación posible se relaciona con la forma en que se revisa el inventario, ya sea en forma continua o periódica. Cuando se aplica un sistema de **revisión continua**, se hace un pedido en el momento en que el inventario baja del punto de reorden especificado. En la **revisión periódica** se verifica el nivel del inventario en intervalos discretos, por ejemplo, al final de cada semana, y sólo en estos momentos se toman las decisiones para ordenar, aun cuando el nivel del inventario hubiera bajado del punto de reorden entre los tiempos de revisión. (En la práctica, se puede usar un modelo de revisión periódica para aproximar una política de revisión continua si se toman intervalos suficientemente pequeños.)

18.3 MODELOS DETERMINÍSTICOS DE REVISIÓN CONTINUA

La situación de inventarios más común que enfrentan los fabricantes, distribuidores y comerciantes es que los niveles de inventarios se reducen con el tiempo y después se reabastecen con la llegada de nuevas unidades. Una representación de esta situación es el **modelo del lote económico** o **modelo EOQ** (*economic order quantity*).

Se supone que los artículos bajo consideración se sacarán en forma continua a una *tasa constante conocida* denotada por d; es decir, la demanda es de d unidades por unidad de tiempo. También se supone que el inventario se reabastece (al producir u ordenar) un lote de tamaño fijo (Q unidades), donde las Q unidades llegan juntas en el tiempo deseado. En el caso del *modelo EOQ básico* que se presentará primero, los únicos costos que se consideran son

K = costo de preparación para ordenar un lote,
c = costo unitario de producir o comprar cada unidad,
h = costo de mantener el inventario por unidad, por unidad de tiempo.

El objetivo consiste en determinar con qué frecuencia y en qué cantidad se debe reabastecer el inventario de manera que se minimice la suma de estos costos por unidad de tiempo.

Se supondrá un proceso de *revisión continua*, por lo que el inventario se puede reabastecer cuando el nivel baje lo suficiente. Primero se supondrá que no se admiten faltantes (pero después se relajará este supuesto). Con la tasa de demanda fija, se pueden evitar los faltantes al reabastecer el inventario cada vez que el nivel baje a cero, enfoque que también minimiza el costo de mantener. En la figura 18.1 se describe el patrón de los niveles de inventario que resulta al comenzar en el tiempo 0 si se produce u ordena un lote de Q unidades, con el fin de aumentar el inventario inicial de 0 a Q y repetir el proceso cada vez que el inventario desciende hasta 0.

El ejemplo 1 de la sección 18.1 (fabricación de bocinas para televisores) se ajusta a este modelo, por lo cual se usará para ilustrar el siguiente análisis.

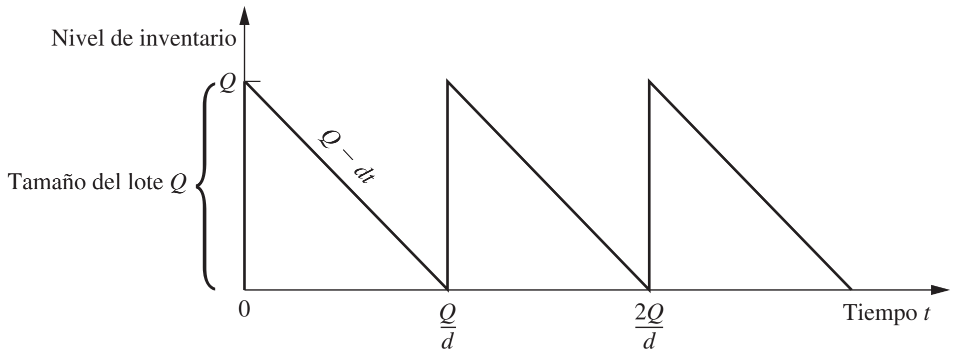
Modelo EOQ básico

Para resumir, además de los costos especificados, el modelo EOQ básico tiene los siguientes supuestos.

Supuestos (modelo EOQ básico).

1. Se conoce la *tasa de demanda* de d unidades por unidad de tiempo.
2. La cantidad ordenada (Q) para reabastecer el inventario llega de una sola vez cuando se desea, es decir, cuando el nivel del inventario baja hasta 0.
3. No se permiten faltantes.

En cuanto al supuesto 2, es común que transcurra un lapso desde que se coloca una orden hasta el momento en que se recibe. Como se indica en la sección 18.2, el tiempo entre colocar una orden



■ **FIGURA 18.1**
Diagrama del nivel de inventario como función del tiempo del modelo EOQ básico.

y recibirla se conoce como *tiempo de entrega*. El nivel de inventario en el que se coloca la orden se llama **punto de reorden**. Para satisfacer el supuesto 2, este punto de reorden debe establecerse como

Punto de reorden = (tasa de demanda) × (tiempo de entrega).

De esta forma, el supuesto 2 asume de manera implícita un tiempo de entrega *constante*.

El tiempo entre reabastecimientos consecutivos del inventario (los segmentos de recta verticales de la figura 18.1) se conoce como *ciclo*. En el ejemplo de las bocinas, un ciclo puede interpretarse como el tiempo que pasa entre las corridas de producción. Si se producen 24 000 bocinas en cada corrida y después se usan a una tasa de 8 000 por mes, la longitud del ciclo es de 24 000/8 000 = 3 meses. En general, la longitud del ciclo es Q/d.

El costo total por unidad de tiempo T se obtiene a partir de los siguientes componentes.

Costo de producir u ordenar por ciclo = $K + cQ$.

El nivel de inventario promedio durante un ciclo es $(Q + 0)/2 = Q/2$ unidades, y el costo correspondiente es $hQ/2$ por unidad de tiempo. Como la longitud del ciclo es Q/d,

$$\text{Costo de mantener inventario por ciclo} = \frac{hQ^2}{2d}.$$

Por lo tanto,

$$\text{Costo total por ciclo} = K + cQ + \frac{hQ^2}{2d},$$

por lo que el costo total por unidad de tiempo es

$$T = \frac{K + cQ + hQ^2/(2d)}{Q/d} = \frac{dK}{Q} + dc + \frac{hQ}{2}.$$

El valor de Q que minimiza T (Q^*), se encuentra al establecer la primera derivada igual a cero (y al observar que la segunda derivada es positiva), de donde se obtiene

$$-\frac{dK}{Q^2} + \frac{h}{2} = 0,$$

de manera que

$$Q^* = \sqrt{\frac{2dK}{h}},$$

que es la bien conocida *fórmula EOQ*.[2] (Algunas veces también se conoce como *fórmula de la raíz cuadrada*.) El *tiempo de ciclo* correspondiente, sea t^*, es

$$t^* = \frac{Q^*}{d} = \sqrt{\frac{2K}{dh}}.$$

Es interesante observar que Q^* y t^* cambian de maneras que son aceptables intuitivamente cuando se hace un cambio en K, h o d. Cuando el costo fijo K crece, tanto Q^* como t^* crecen (menos preparaciones). Si el costo unitario de mantener h aumenta, tanto Q^* como t^* disminuyen (niveles de inventario menores). A medida que la tasa de demanda d crece, Q^* también lo hace (lotes más grandes), pero t^* disminuye (preparaciones más frecuentes).

Se aplicarán estas fórmulas de Q^* y t^* en el ejemplo de las bocinas. Los valores apropiados de los parámetros dados en la sección 18.1 son

$$K = 12\ 000, \qquad h = 0.30, \qquad d = 8\ 000,$$

de manera que

$$Q^* = \sqrt{\frac{(2)(8\ 000)(12\ 000)}{0.30}} = 25\ 298$$

y

$$t^* = \frac{25\ 298}{8\ 000} = 3.2 \text{ meses.}$$

En consecuencia, la solución óptima es hacer una preparación de la línea de producción de bocinas cada 3.2 meses y producir 25 298 bocinas cada vez. (La curva de costo es bastante plana cerca del valor óptimo, por lo que cualquier cantidad de producción similar que sea más conveniente, como 24 000 bocinas cada 3 meses, sería muy cercana a la óptima.)

En la sección Worked Examples del sitio en internet de este libro se incluye **otro ejemplo** de aplicación del modelo EOQ básico, pero en ese caso, además, es necesario realizar un análisis de sensibilidad considerable.

Modelo EOQ con faltantes planeados

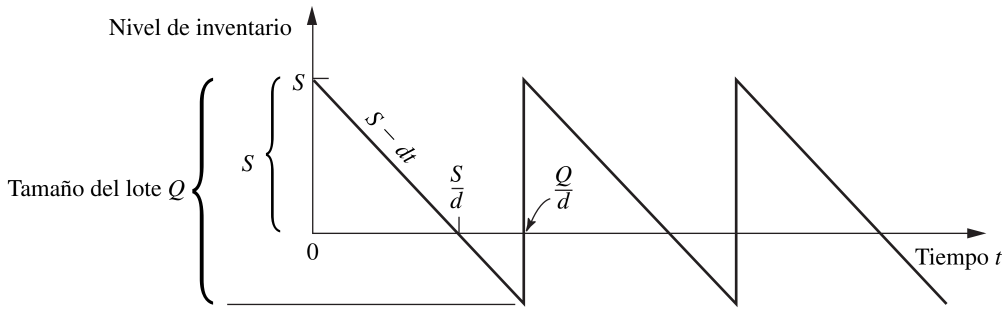
Uno de los inconvenientes en la administración de cualquier sistema de inventarios es que ocurran faltantes (también llamados *órdenes pendientes*), que no es otra cosa que la demanda que no se satisface debido a que el inventario se ha agotado. Esta situación causa muchos dolores de cabeza, que incluyen tratar con clientes molestos y realizar el trabajo adicional de registros para cumplir esa demanda más tarde (*se permiten faltantes*) al reabastecer el inventario. Si se supone que no se permite planear que ocurran faltantes, el modelo básico EOQ que se presentó satisface el deseo común de los administradores para evitar los faltantes lo más posible. (De cualquier forma, los faltantes no planeados pueden ocurrir si la tasa de demanda y las entregas no se ajustan a lo programado.)

Sin embargo, existen situaciones limitadas en las que permitir faltantes planeados tiene sentido desde el punto de vista administrativo. El requisito más importante es que los clientes, en general, estén dispuestos a aceptar un retraso razonable en la recepción de sus pedidos si es necesario. Si así es, los costos de incurrir en faltantes descritos en las secciones 18.1 y 18.2 (incluso la pérdida de negocios futuros) no serán exorbitantes. Si el costo de mantener inventarios es alto en relación con los costos de faltantes, bajar el nivel de inventarios y permitir faltantes breves ocasionales puede ser una buena decisión.

El **modelo EOQ con faltantes planeados** toma en cuenta este tipo de situación y sustituye sólo el tercer supuesto del modelo básico EOQ por el siguiente:

> Ahora se permiten faltantes planeados. Cuando ocurre un faltante, los clientes afectados esperan que el producto esté nuevamente disponible. Sus órdenes pendientes se satisfacen de inmediato cuando llega la cantidad ordenada para reabastecer el inventario.

[2] Un registro histórico interesante de esta fórmula y modelo, incluso la reimpresión de un artículo de 1913 que dio inicio a todo este desarrollo, puede encontrarse en D. Erlenkotter, “Ford Whitman Harris and the Economic Order Quantity Model”, en *Operations Research*, **38**: 937-950, 1990.



■ **FIGURA 18.2**
Diagrama del nivel de inventario como una función del tiempo en el modelo EOQ con faltantes planeados.

Bajo estos supuestos, el patrón de niveles de inventario en el tiempo tiene la apariencia que se muestra en la figura 18.2. El aspecto de dientes de sierra es el mismo que el de la figura 18.1. Sin embargo, en este caso los niveles de inventario se extienden a valores negativos que reflejan el número de unidades del producto que faltaron o que están pendientes de entregar.

Sea

p = costo de faltantes por unidad que falta por unidad de tiempo que falta,
S = nivel de inventario justo después de recibir un lote de Q unidades,
$Q - S$ = faltante en inventario justo antes de recibir un lote de Q unidades.

El costo total por unidad de tiempo se obtiene a partir de los siguientes componentes:

Costo de producir u ordenar por ciclo $= K + cQ$.

Durante cada ciclo, el nivel de inventario es positivo durante un tiempo S/d. El nivel del inventario promedio *durante este tiempo* es $(S + 0)/2 = S/2$ artículos por unidad de tiempo y el costo correspondiente es $hS/2$ por unidad de tiempo. Entonces,

$$\text{Costo de mantener el inventario por ciclo} = \frac{hS}{2}\frac{S}{d} = \frac{hS^2}{2d}.$$

De manera similar, los faltantes ocurren durante un tiempo $(Q - S)/d$. La cantidad promedio de faltantes *durante este tiempo* es $(0 + Q - S)/2 = (Q - S)/2$ artículos, y el costo correspondiente es $p(Q - S)/2$ por unidad de tiempo. Así,

$$\text{Costo de faltantes por ciclo} = \frac{p(Q-S)}{2}\frac{Q-S}{d} = \frac{p(Q-S)^2}{2d}.$$

Por lo tanto,

$$\text{Costo total por ciclo} = K + cQ + \frac{hS^2}{2d} + \frac{p(Q-S)^2}{2d},$$

y el *costo total por unidad de tiempo* es

$$T = \frac{K + cQ + hS^2/(2d) + p(Q-S)^2/(2d)}{Q/d}$$

$$= \frac{dK}{Q} + dc + \frac{hS^2}{2Q} + \frac{p(Q-S)^2}{2Q}.$$

En este modelo hay dos variables de decisión (S y Q) y los valores óptimos (S^* y Q^*) se encuentran al igualar a cero las derivadas parciales $\partial T/\partial S$ y $\partial T/\partial Q$. Entonces,

$$\frac{\partial T}{\partial S} = \frac{hS}{Q} - \frac{p(Q-S)}{Q} = 0.$$

$$\frac{\partial T}{\partial Q} = -\frac{dK}{Q^2} - \frac{hS^2}{2Q^2} + \frac{p(Q-S)}{Q} - \frac{p(Q-S)^2}{2Q^2} = 0.$$

Al resolver estas ecuaciones en forma simultánea se obtiene

$$S^* = \sqrt{\frac{2dK}{h}}\sqrt{\frac{p}{p+h}}, \qquad Q^* = \sqrt{\frac{2dK}{h}}\sqrt{\frac{p+h}{p}}.$$

La longitud óptima del ciclo t^* está dada por

$$t^* = \frac{Q^*}{d} = \sqrt{\frac{2K}{dh}}\sqrt{\frac{p+h}{p}}.$$

El faltante máximo es

$$Q^* - S^* = \sqrt{\frac{2dK}{p}}\sqrt{\frac{h}{p+h}}.$$

Además, en la figura 18.2 se observa que la fracción de tiempo en que no existen faltantes es

$$\frac{S^*/d}{Q^*/d} = \frac{p}{p+h},$$

que es independiente de K.

Cuando el valor de p o de h se hace mucho más grande que el otro, las cantidades anteriores se comportan de manera intuitiva. En particular, cuando $p \to \infty$ con h constante (los costos por faltantes dominan), $Q^* - S^* \to 0$ mientras que tanto Q^* como t^* convergen a sus valores dados en el modelo EOQ básico. Aunque el modelo actual permite faltantes, $p \to \infty$ implica que no vale la pena tenerlos.

Por otro lado, cuando $h \to \infty$ con p constante (de manera que dominan los costos de mantener inventario), $S^* \to 0$. Por ello, tener $h \to \infty$ hace que no sea económico tener niveles de inventario positivos, con lo que cada nuevo lote de Q^* unidades no va más allá de eliminar los actuales faltantes de inventario.

Si en el ejemplo de las bocinas se permiten faltantes planeados, el *costo por faltantes* se estimó en la sección 18.1 como

$$p = 1.10.$$

De nuevo,

$$K = 12\,000, \qquad h = 0.30, \qquad d = 8\,000,$$

por lo que ahora

$$S^* = \sqrt{\frac{(2)(8\,000)(12\,000)}{0.30}}\sqrt{\frac{1.1}{1.1+0.3}} = 22\,424,$$

$$Q^* = \sqrt{\frac{(2)(8\,000)(12\,000)}{0.30}}\sqrt{\frac{1.1+0.3}{1.1}} = 28\,540,$$

y

$$t^* = \frac{28\,540}{8\,000} = 3.6 \text{ meses.}$$

En consecuencia, la línea de producción de bocinas debe prepararse cada 3.6 meses para producir 28 540 unidades. El faltante máximo que se permite es de 6 116 bocinas. Note que Q^* y t^* no difieren mucho de los valores del caso en que no se permiten faltantes. La razón es que p es mucho mayor que h.

Modelo EOQ con descuentos por cantidad

Cuando se especificaron las componentes de costos, los modelos anteriores suponen que el costo por unidad de un artículo es el mismo sin importar la cantidad que compone el lote. En realidad,

este supuesto da como resultado que las soluciones óptimas sean independientes del costo unitario. El *modelo EOQ con descuentos por cantidad* sustituye ese supuesto con el siguiente.

> El costo unitario de un artículo depende de la cantidad de unidades que integren el lote. En particular, se proporciona un incentivo para colocar una orden grande al cambiar el costo unitario de cantidades pequeñas por un costo unitario menor en lotes más grandes y quizá un costo unitario todavía más pequeño para lotes aún más grandes.

De otra manera, los supuestos son los mismos que los del modelo EOQ básico.

Para ilustrar este modelo, considere el ejemplo de las bocinas para TV de la sección 18.1. Suponga que el costo unitario de *cada* bocina es $c_1 = \$11$ si se producen menos de 10 000, $c_2 = \$10$ si la producción está entre 10 000 y 80 000 bocinas y $c_3 = \$9.50$ si la producción es mayor a 80 000 unidades. ¿Cuál es la política óptima? La solución al problema específico revelará la metodología general.

A partir de los resultados del modelo básico EOQ, el costo total por unidad de tiempo T_j, si el costo de producción es c_j, está dado por

$$T_j = \frac{dK}{Q} + dc_j + \frac{hQ}{2}, \qquad \text{para } j = 1, 2, 3.$$

(En esta expresión se supone que h es independiente del costo unitario de los artículos, pero un pequeño refinamiento común sería hacer h proporcional al costo unitario para reflejar el hecho de que el costo de capital comprometido en el inventario varía de esta manera.) En la figura 18.3 se muestra una gráfica de T_j contra Q, para cada j, donde la parte sólida de cada curva se extiende en el intervalo de valores factibles de Q de esa categoría de descuento.

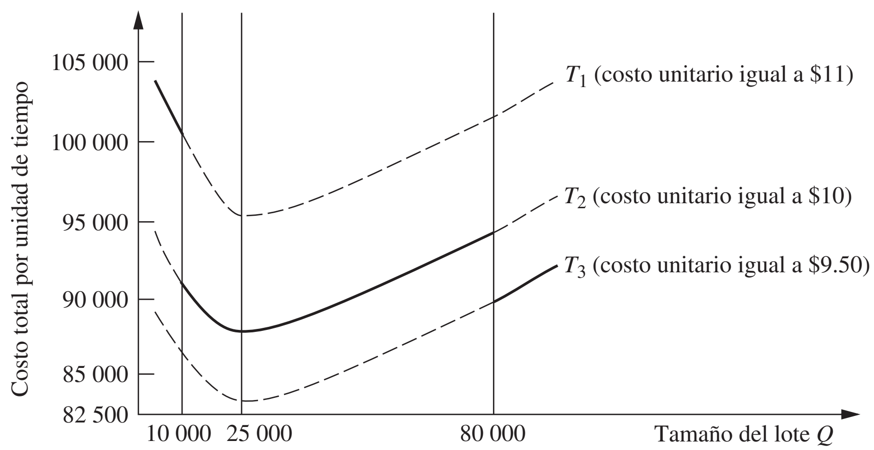
En el caso de cada curva, el valor de Q que minimiza T_j se encuentra igual que en el modelo básico del lote económico. Para $K = 12\,000$, $h = 0.30$ y $d = 8\,000$, este valor es

$$\sqrt{\frac{(2)(8\,000)(12\,000)}{0.30}} = 25\,298.$$

(Si h no fuera independiente del costo unitario de los artículos, el valor de Q que minimiza sería un poco diferente para cada curva.) Este valor de Q que minimiza es un valor factible para la función de costo T_2. Para cualquier Q fija, $T_2 < T_1$, de manera que T_1 se puede eliminar. Sin embargo, T_3 no puede eliminarse de forma directa. Su valor factible mínimo (que ocurre en $Q = 80\,000$) debe compararse con T_2 evaluado en 25 298 (que es 87 589 dólares). Como T_3 evaluado en 80 000 es igual a 89 200 dólares, es mejor producir cantidades de 25 298, por lo cual esta cantidad es el valor óptimo de este conjunto de descuentos por cantidad.

Si la cantidad descontada hubiera llevado a un costo unitario de 9 dólares (en lugar de $9.50) cuando la producción excediera de 80 000 bocinas, entonces T_3 evaluado en 80 000 hubiera sido igual a 85 200 dólares y la cantidad óptima a producir sería de 80 000 unidades.

FIGURA 18.3
Costo total por unidad de tiempo en el ejemplo de las bocinas con descuentos por cantidad.



Aunque este análisis concierne a un problema muy específico, el mismo enfoque se puede aplicar a cualquier problema similar. A continuación se presenta un resumen del procedimiento general.

1. En el caso de cada costo unitario disponible c_j, use la fórmula del modelo EOQ para calcular la cantidad óptima de ordenar Q_j^*.
2. En el caso de cada c_j donde Q_j^* se encuentra dentro del intervalo factible de cantidades por ordenar para c_j, calcule el costo total correspondiente por unidad de tiempo T_j.
3. En el caso de cada c_j donde Q_j^* no está dentro del intervalo factible, determine la cantidad por ordenar Q_j que se encuentra en el punto terminal más cercano a Q_j^*. Calcule el costo total por unidad de tiempo T_j para Q_j y c_j.
4. Compare las T_j que obtuvo para todas las c_j y elija la T_j mínima. Después seleccione la cantidad por ordenar Q_j que obtuvo en los pasos 2 o 3 que proporciona esta T_j mínima.

Se puede usar un análisis parecido para manejar otros tipos de descuentos por cantidad, como los descuentos incrementales por cantidad, donde se incurre en un costo c_0 en los primeros q_0 artículos, c_1 en los siguientes q_1 artículos, y así sucesivamente.

Algunas plantillas útiles de Excel

Para su conveniencia, se incluyeron cinco plantillas de Excel para los modelos EOQ en el archivo de Excel de este capítulo en el sitio en internet de este libro. Dos de las plantillas son para el modelo básico EOQ. En ambos casos, se introducen los datos básicos (d, K y h), lo mismo que los tiempos de entrega y el número de días hábiles por año de la empresa. Después, la plantilla calcula el gasto total anual de la empresa por preparaciones y costos de mantener, y la suma de estos dos costos (el *costo variable total*). También calcula el *punto de reorden*, esto es, el nivel de inventario al que debe colocarse la orden para que el reabastecimiento llegue cuando el nivel de inventario baja hasta 0. Una plantilla (la *versión de Solver*) permite introducir cualquier tamaño de orden que desee y ver cuál sería el costo anual y el punto de reorden. Esta versión también permite usar Excel Solver para obtener la cantidad óptima de la orden. La segunda plantilla (la *versión analítica*) usa la fórmula EOQ para obtener la cantidad óptima de pedido.

También se proporciona el par de plantillas correspondientes al modelo EOQ con faltantes planeados. Después de introducir los datos (incluso el costo unitario por faltantes p), cada plantilla obtiene los diferentes costos anuales (que incluyen el costo anual por faltantes). Con la versión de Solver se pueden probar valores de la cantidad por ordenar Q y el faltante máximo $Q - S$ u obtener los valores óptimos, mientras que la versión analítica usa las fórmulas de Q^* y $Q^* - S^*$ para obtener los valores óptimos. También se incluye en los resultados el nivel de inventario máximo S^* correspondiente.

La plantilla final es una versión analítica del modelo EOQ con descuentos por cantidad. Incluye el refinamiento de que el costo unitario de mantener h es proporcional al costo unitario c, de manera que

$$h = Ic,$$

donde el factor de proporcionalidad I se conoce como *tasa de costo de mantener inventario*. En consecuencia, los datos incluyen I junto con d y K. También debe introducirse el número de categorías de descuento (donde la categoría de la menor cantidad sin descuento es una de ellas) y el precio unitario y el intervalo de cantidades por ordenar para cada categoría. La plantilla encuentra el tamaño de la orden factible que minimiza el costo total anual de cada categoría y muestra los costos anuales individuales (junto con el costo de compra anual) que se obtiene. Con esta información, identifica la cantidad óptima por ordenar global y el costo total anual que resulta.

Todas estas plantillas pueden ser útiles para calcular con rapidez una gran cantidad de información después de introducir los datos básicos del problema. Sin embargo, tal vez el uso más importante sea realizar un análisis de sensibilidad sobre estos datos. Es sencillo ver de inmediato cómo cambian los resultados con cualquier cambio específico en los datos al colocar los nuevos valores en la hoja de cálculo. Hacer esto varias veces para diferentes cambios en los datos es una manera de realizar el análisis de sensibilidad.

Observaciones sobre los modelos EOQ

1. Si se supone que el costo unitario de un artículo es constante en el tiempo, sin importar el tamaño del lote (como en los dos primeros modelos EOQ), este costo unitario no aparece en la solución óptima del tamaño del lote. Este resultado ocurre porque no importa qué política se use, se requiere el mismo número de unidades y así este costo por unidad de tiempo es fijo.
2. El análisis de los modelos EOQ supone que el tamaño del lote Q es constante de un ciclo a otro. El tamaño del lote *óptimo* Q^* en realidad minimiza el costo total por unidad de tiempo de cualquier ciclo, por lo que el análisis muestra que debe usarse este tamaño de lote constante de un ciclo a otro aunque no se haya supuesto un tamaño constante.
3. En estos modelos, el nivel del inventario óptimo en el que debe reabastecerse nunca puede ser mayor que cero. Si se espera hasta que el inventario baje a cero (o a menos de cero cuando se permiten faltantes) se reducen tanto los costos de mantener como la frecuencia con la que se incurre en el costo fijo K. Sin embargo, si no se cumplen por completo los supuestos de *una tasa de demanda constante y conocida* y de que *la cantidad ordenada llega justo cuando se desea* (debido al tiempo de entrega constante), puede ser prudente planear un "inventario de seguridad" que queda cuando está programado reabastecer el inventario. Este objetivo se logra si se coloca el punto de reorden arriba del que implica el modelo.
4. Los supuestos básicos de los modelos EOQ son exigentes. Sin embargo, en la práctica rara vez se satisfacen por completo. Por ejemplo, aun cuando se planea una tasa de demanda constante (como en la producción de bocinas de la sección 18.1), es posible que ocurran interrupciones y variaciones en esa tasa. También es difícil satisfacer el supuesto de que la orden para reabastecer llega justo cuando se desea. Aunque la programación dice que el tiempo de entrega es constante, existen variaciones en los tiempos reales. Por fortuna, se ha comprobado que los modelos EOQ son creíbles pues lo normal es que proporcionen resultados muy cercanos al óptimo incluso cuando sus supuestos sean una aproximación a la realidad. Ésta es una razón clave del amplio uso de estos modelos en la práctica. Sin embargo, en los casos en que los supuestos se violan de manera significativa, es importante hacer un análisis preliminar para evaluar si el modelo EOQ es adecuado antes de usarlo. Este análisis preliminar debe centrarse en calcular el costo total por unidad de tiempo que proporciona el modelo para diferentes tamaños de orden, y después en evaluar cómo cambia la curva de este costo con supuestos más realistas.

Diferentes tipos de demanda de un producto

El ejemplo 2 (distribuidor de bicicletas) que se introdujo en la sección 18.1 se centró en la administración del inventario de un modelo de bicicleta. La demanda de este producto la generan los clientes del distribuidor (comerciantes) que las compran para reabastecer sus inventarios de acuerdo a su propia programación. El distribuidor no tiene control sobre esta demanda. Debido a que la venta de este modelo es independiente de los otros, su demanda, que ni siquiera depende de la de los otros productos de la compañía, recibe el nombre de **demanda independiente**.

La situación es diferente a la de las bocinas del ejemplo de la sección 18.1. Aquí, el producto bajo consideración, bocinas para televisores, es sólo una componente que se ensambla en el producto final de la compañía, los televisores. En consecuencia, la demanda de bocinas depende de la demanda de televisores. El patrón de esta demanda de bocinas está determinado de manera interna por el programa de producción que establece la compañía y se ajusta a la tasa de producción de la línea que produce los televisores. Esta demanda se conoce como **demanda dependiente**.

La compañía fabricante de televisores produce un número considerable de productos, como las partes y subensambles, que se convierten en componentes de los televisores. Igual que las bocinas, estos productos también son **productos de demanda dependiente**.

Debido a las dependencias e interrelaciones involucradas, la administración de inventarios de productos de demanda dependiente puede ser mucho más complicada que en el caso de demanda independiente. Una técnica popular para ayudar en esta tarea es el sistema de **planeación de los requerimientos de materiales**, abreviado **MRP** (*material requirements planning*). El MRP es un sistema basado en computadora para planear, programar y controlar la producción de todos los componentes de un producto final. El sistema comienza por "explotar" el producto pues lo divide en subensambles y después en todos sus componentes individuales. Se desarrolla un programa de producción que utiliza la demanda y el tiempo de entrega de cada componente para determinar la

demanda y el tiempo de entrega de los componentes subsecuentes en el proceso. Además de un *programa maestro de producción* del producto final, una *lista de materiales* proporciona información detallada de todos los componentes. Los registros del estado del inventario proporcionan los niveles actuales, el número de unidades ordenadas, etc., de todos los componentes. Cuando deben ordenarse más unidades de un componente, el sistema MRP genera en forma automática una orden de compra dirigida al proveedor o una orden de trabajo para el departamento interno que produce el componente.[3]

Papel de la administración de inventarios justo a tiempo (JIT)

Cuando se usó el modelo EOQ básico para calcular el lote óptimo de producción en el ejemplo de las bocinas, se obtuvo una cantidad muy grande (25 298 bocinas), lo cual implica pocas preparaciones para iniciar las corridas de producción (sólo una vez cada 3.2 meses). Por otro lado, significa la existencia de niveles promedio de inventario muy altos (12 649 bocinas), lo que conduce a un costo total anual de mantenimiento de más de 45 000 dólares.

La razón básica de este costo es el costo de preparación tan alto $K = \$12\ 000$ de cada corrida de producción. El costo de preparación es tan grande porque las instalaciones deben empezar de cero cada vez. En consecuencia, aun con menos de cuatro corridas de producción por año, el costo anual de preparación supera los 45 000 dólares, igual que los costos de mantener.

En lugar de seguir tolerando un costo de preparación de 12 000 dólares cada vez en el futuro, otra opción es buscar la manera de reducir este costo. Una posibilidad es desarrollar métodos para transferir con rapidez las máquinas de un uso a otro. Otra es dedicar un grupo de máquinas a la producción de bocinas para que permanezcan listas para ello entre corridas de producción, esto es, preparadas para iniciar otra corrida cuando sea necesario.

Suponga que el costo de preparación se puede reducir en forma drástica desde 12 000 dólares hasta $K = \$120$, lo cual abatiría el tamaño del lote óptimo de producción de 25 298 bocinas a $Q^* = 2\ 530$, de manera que 3 veces por mes se realizaría una pequeña corrida más. De manera simultánea se reduciría tanto el costo anual de preparación como el costo anual de mantener inventarios de más desde 45 000 hasta sólo 4 500 dólares cada uno. Debido a pequeñas pero frecuentes corridas (poco costosas), las bocinas se producirían *justo a tiempo* para su ensamble en los televisores.

En realidad, *justo a tiempo* es una filosofía correctamente desarrollada para administrar inventarios. Un sistema de inventarios **justo a tiempo** (JIT, *just-in-time*) hace hincapié en la reducción de los niveles de inventarios hasta el mínimo, y en proporcionar los artículos justo a tiempo a medida que se necesite. A este enfoque, que se desarrolló en Japón, primero en la Toyota Company a finales de la década de 1950, se le atribuye parte del crédito por los asombrosos aumentos de la productividad japonesa en la última parte del siglo XX. Este enfoque también se popularizó en otras partes del mundo, incluso en Estados Unidos en años más recientes.[4]

Aunque en ocasiones la filosofía de justo a tiempo se malinterpreta como incompatible con el uso del modelo EOQ (pues éste da una orden grande cuando el costo de preparación es alto), en realidad son herramientas complementarias. Un sistema de inventarios JIT se centra en encontrar formas de reducir de manera drástica los costos de preparación para que la cantidad óptima de la orden sea pequeña. Tal sistema también trata de encontrar la manera de reducir el tiempo de entrega de la recepción de una orden pues así se reduce la incertidumbre acerca del número de unidades que serán necesarias en el momento en que ocurra la entrega. Otro punto importante es la mejora del mantenimiento preventivo para que las instalaciones de producción requeridas estén disponibles para producir las unidades cuando son necesarias. Otro aspecto de gran importancia es la mejora del proceso de producción para garantizar la buena calidad. Contar el número correcto de unidades justo a tiempo no equivale a tener una vía libre para incluir unidades defectuosas.

[3] Una serie de artículos en pp. 32-44 del número de septiembre de 1996 de *IIE Solutions* proporciona mayor información acerca del MRP.

[4] Si desea más información acerca de las aplicaciones del sistema JIT en Estados Unidos, vea R. E. White, J. N. Pearson y J. R. Wilson, “JIT Manufacturing: A Survey of Implementations in Small and Large U.S. Manufacturing”, en *Management Science*, **45**: 1-15, 1999. También consulte H. Chen, M. Z. Frank y O. Q. Wu, “What Actually Happened to the Inventories of American Companies Between 1981 y 2000”, en *Management Science*, **51**(7): 1015-1031, julio de 2005.

En términos más generales, el enfoque de la filosofía justo a tiempo se concentra en *evitar cualquier forma de desperdicio* que pueda ocurrir en el proceso de producción. Una de ellas es el inventario innecesario, otras son los costos de preparación innecesariamente altos, los tiempos de entrega demasiado largos, las instalaciones de producción poco operativas cuando se requieren y los artículos defectuosos. Minimizar estas formas de desperdicio es un componente clave de la administración de inventarios de nivel superior.

18.4 MODELO DETERMINÍSTICO CON REVISIÓN PERIÓDICA

El análisis de la sección anterior exploró el modelo del lote económico (EOQ). Los resultados obtenidos se basan en el supuesto de que la tasa de demanda es constante. Cuando este supuesto se relaja, es decir, cuando se permite que varíen las cantidades que deben retirarse del inventario de un periodo a otro, la *fórmula EOQ* ya no asegura una solución de costo mínimo.

Considere el siguiente modelo de revisión periódica. Deben planearse cuánto producir u ordenar (si es necesario) los siguientes n periodos para reabastecer el inventario al principio de cada uno de éstos. (La orden de reabastecer el inventario puede requerir la *compra* de las unidades o su *producción*; esto último es lo más común cuando se aplica el presente modelo, por lo que se usará el término *producir* las unidades.) Las demandas en los respectivos periodos son *conocidas* (pero *no* son las mismas en todos los periodos) y se denotan por

r_i = demanda en el periodo i, para $i = 1, 2, \ldots, n$.

Estas demandas se deben satisfacer a tiempo. No se tiene un inventario inicial, pero hay tiempo para hacer una entrega al principio del periodo 1.

Los costos incluidos en este modelo son similares a los del primer modelo EOQ:

K = costo de preparación para producir u ordenar artículos para reabastecer el inventario al inicio del periodo,
c = costo unitario de producir u ordenar cada artículo,
h = costo de mantener en inventario cada artículo que queda al final del periodo.

Observe que este costo de mantener, h, se evalúa sólo con respecto al inventario que queda al final del periodo. También existen costos de mantener las unidades en inventario durante una parte del periodo antes de extraerlas para satisfacer la demanda. Sin embargo, éstos son *costos fijos* independientes de la política de inventarios y no son relevantes para el análisis. Sólo los costos *variables* a los que afecta la política elegida, como los costos de mantener adicionales en que se incurre cuando se tienen unidades en inventario de un periodo al siguiente son relevantes para elegir la política de inventarios.

Según el mismo razonamiento, el costo unitario c es un costo fijo irrelevante porque, durante todos los periodos, todas las políticas de inventarios producen el mismo número de unidades al mismo costo. Por lo tanto, en adelante c se eliminará del análisis.

El objetivo es minimizar el costo total durante los n periodos. Esto se logra si se pasan por alto los costos fijos y, como se ilustra en el siguiente ejemplo, se minimiza el costo total variable de n periodos.

Un ejemplo

Un fabricante de aeronaves se especializa en la producción de aviones pequeños. Acaba de recibir un pedido de una gran corporación de 10 aviones jet ejecutivos especiales para uso de la alta administración de la empresa. La orden pide que se entreguen tres aviones (que se pagarán) durante los meses del próximo invierno (periodo 1), dos más en la primavera (periodo 2), tres en el verano (periodo 3) y los últimos dos durante el otoño (periodo 4).

La preparación de las instalaciones de producción para cumplir con las especificaciones de la corporación para fabricar estos aviones implica un costo de 2 millones de dólares. El fabricante tiene la capacidad de producir los 10 aviones en un par de meses, cuando la temporada de invierno haya comenzado. Sin embargo, esto significaría mantener siete de ellos en inventario, a un costo de 200 000 dólares por avión por periodo, hasta la fecha de entrega programada. Para reducir o eli-

minar estos costos sustanciales de mantener el inventario, tal vez valga la pena producir un número menor de aviones ahora y después repetir la preparación (aunque se incurra de nuevo en el costo de 2 millones de dólares) en algunos o todos los periodos subsecuentes con corridas de producción pequeñas. La administración desea determinar el programa de producción menos costoso para satisfacer esta orden.

Entonces, con base en la notación del modelo, las demandas de este avión en particular durante los cuatro periodos siguientes (temporadas) son

$$r_1 = 3, \qquad r_2 = 2, \qquad r_3 = 3, \qquad r_4 = 2.$$

Si se usan unidades de millones de dólares, el costo relevante es

$$K = 2, \qquad h = 0.2.$$

El problema es determinar cuántos aviones se deben producir (si se producen) para el inicio de cada uno de los cuatro periodos con el fin de minimizar el costo variable total.

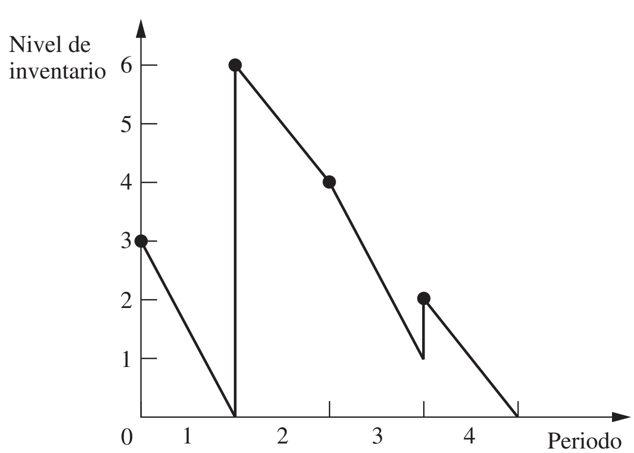
El costo fijo K tan alto es un fuerte incentivo para no producir aviones cada periodo y de preferencia hacerlo sólo una vez. Sin embargo, el significativo costo de mantener h hace poco deseable tener un inventario alto si se produce la demanda completa de los cuatro periodos (10 aviones) al principio. Quizá el mejor enfoque sea una estrategia intermedia que permita producir aviones más de una vez pero menos de cuatro. Por ejemplo, una solución factible (pero no óptima) se describe en la figura 18.4, con la evolución del nivel de inventario durante el siguiente año: estos niveles se obtienen cuando se producen 3 aviones al principio del primer periodo, 6 al principio del segundo y 1 al principio del cuarto. Los puntos indican el nivel de inventario después de cualquier corrida de producción al principio de los cuatro periodos.

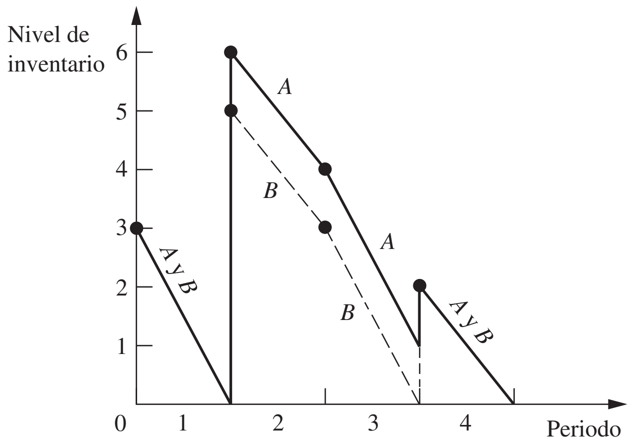
¿Cómo se puede determinar el programa de producción óptimo? En el caso de este modelo en general, la producción (o el surtido de pedidos) es automática en el periodo 1, pero se debe decidir si producir o no en los otros $n - 1$ periodos. Por lo tanto, un enfoque para resolver este modelo es enumerar, en el caso de cada una de las 2^{n-1} combinaciones de decisiones de producción, las cantidades posibles que se pueden producir en cada periodo en que la producción deba ocurrir. Este enfoque es bastante lento y tedioso incluso cuando el tamaño de n es moderado, por lo que es deseable contar con un método más eficiente. Ese método se describirá a continuación en términos generales, y después se regresará a él hasta encontrar el programa de producción óptimo para el ejemplo. Aunque se puede usar el método general cuando se produce o se compra para reabastecer el inventario, con el fin de ser concretos ahora sólo se utilizará la terminología de producción.

Un algoritmo

La clave para desarrollar un algoritmo eficiente para encontrar una *política óptima de inventarios* (o de manera equivalente, un *programa de producción óptimo*) para el modelo anterior está basada en la siguiente observación sobre la naturaleza de una política óptima.

FIGURA 18.4
Niveles de inventario que resultan de un programa de producción muestra en el ejemplo de los aviones.





■ **FIGURA 18.5**
Comparación de dos políticas de inventarios (programas de producción) en el ejemplo de los aviones.

Una política óptima (programa de producción) produce *sólo* cuando el nivel de inventario es *cero*.

Para ilustrar por qué este resultado es cierto, considere para el ejemplo la política que se muestra en la figura 18.4. (Sea la política A.) La política A viola la caracterización anterior de una política óptima porque la producción ocurre al principio del periodo 4 cuando el nivel de inventario es *mayor que cero* (es decir, un avión). Sin embargo, es muy sencillo ajustar esta política para satisfacer la caracterización anterior con sólo producir una unidad menos en el periodo 2 y una más en el 4. Esta política ajustada (sea B) se muestra mediante la línea punteada de la figura 18.5 siempre que B difiera de A (la línea continua). Ahora observe que la política B *debe* tener un costo total menor que la política A. Los costos de preparación y de producción de ambas políticas son los mismos. Sin embargo, el costo de mantener el inventario es más pequeño en el caso de B que en el de A porque B tiene un inventario menor en los periodos 2 y 3 (y el mismo inventario en los otros periodos). Por lo tanto, B es mejor que A, y entonces A no puede ser óptima.

Esta caracterización de las políticas óptimas se puede usar para identificar las políticas que no son óptimas. Además, como este enfoque implica que las únicas opciones para la cantidad producida al principio del periodo i son $0, r_i, r_i + r_{i+1}, \ldots,$ o $r_i + r_{i+1} + \ldots + r_n$ se puede explotar para obtener un algoritmo eficiente relacionado con el enfoque de *programación dinámica determinística* que se describió en la sección 10.3.

En particular, defina

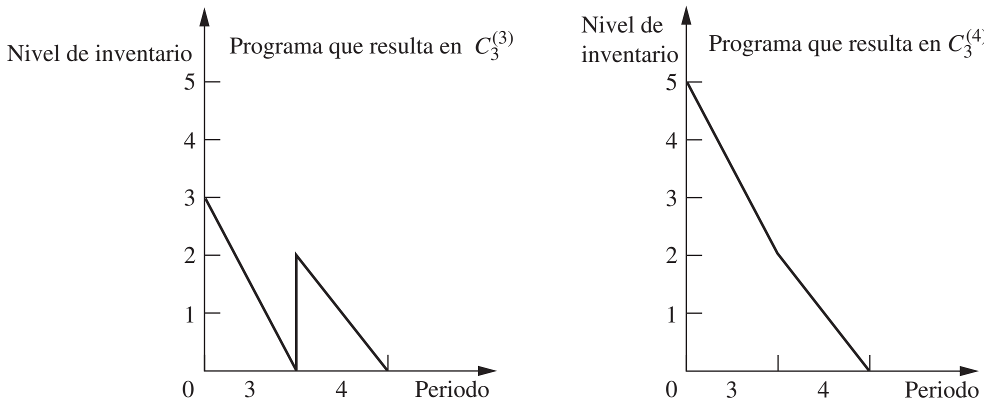
C_i = costo total de una política óptima para los periodos $i, i + 1, \ldots, n$ cuando el periodo i se inicia con inventario cero (antes de producir), para $i = 1, 2, \ldots, n$.

Al usar el enfoque de programación dinámica y resolver *hacia atrás* periodo por periodo, los valores de C_i se pueden encontrar mediante el cálculo primero de C_n, después de C_{n-1}, etc. Así, una vez calculados $C_n, C_{n-1}, \ldots, C_{i+1}$, se puede encontrar C_i a partir de la *relación recursiva*

$$C_i = \underset{j=i,\, i+1,\, \ldots,\, n}{\text{mínimo}} \{C_{j+1} + K + h[r_{i+1} + 2r_{i+2} + 3r_{i+3} + \cdots + (j - i)r_j]\},$$

donde j se puede interpretar como un índice que denota el (final del) periodo cuando el inventario llega a cero por primera vez después de producir al principio del periodo i. En el tiempo que transcurre entre periodo i al periodo j, el término con coeficiente h representa el costo total de *mantener* en ese intervalo. Cuando $j = n$, el término $C_{n+1} = 0$. El *valor* de j que *minimiza* indica que si el nivel de inventario en realidad llega a cero al iniciar el periodo i, entonces la producción en el periodo i debe cubrir toda la demanda desde el periodo i hasta este periodo j.

El algoritmo para resolver este modelo consiste en esencia en obtener $C_n, C_{n-1}, \ldots, C_1$ por turno. Para $i = 1$, el valor de j que minimiza indica que la producción en el periodo 1 debe cubrir la demanda hasta el periodo j, y la segunda producción será en el periodo $j + 1$. Para $i = j + 1$, el nuevo valor de j que minimiza señala el tiempo que la segunda producción cubre la demanda, y así hasta el final. Se ilustrará este enfoque con un ejemplo.



■ **FIGURA 18.6**
Programas de producción alternativos cuando se requiere la producción al principio del periodo 3 en el ejemplo de los aviones.

La aplicación de este algoritmo es mucho más rápida que el enfoque completo de programación dinámica.[5] Igual que en ésta, es necesario determinar C_n, C_{n-1}, …, C_2 antes de obtener C_1, pero el número de cálculos es mucho menor y el número de cantidades posibles de producción se reduce en forma significativa.

Aplicación del algoritmo al ejemplo

De nuevo en el ejemplo de los aviones, primero se considera el caso de encontrar C_4, el costo de la política óptima a partir del inicio del periodo 4 al final del horizonte de planeación:

$$C_4 = C_5 + 2 = 0 + 2 = 2.$$

Para calcular C_3 debe considerarse dos casos respecto de la primera vez que el inventario llega a cero después del periodo 3, lo que puede ocurrir 1) al final del tercer periodo o 2) al final del cuarto periodo. En la relación recursiva de C_3 estos dos casos corresponden a 1) $j = 3$ y 2) $j = 4$. Denote los costos correspondientes (el lado derecho de la relación recursiva con esta j) por $C_3^{(3)}$ y $C_3^{(4)}$, respectivamente. La política asociada con $C_3^{(3)}$ dice que se debe producir sólo para el periodo 3 y después seguir la política óptima en el periodo 4, mientras que la política asociada con $C_3^{(4)}$ dice que debe producirse para los periodos 3 y 4. De esta forma, el costo C_3 es el mínimo de $C_3^{(3)}$ y $C_3^{(4)}$. Estos casos se reflejan en las políticas dadas en la figura 18.6.

$$C_3^{(3)} = C_4 + 2 = 2 + 2 = 4.$$
$$C_3^{(4)} = C_5 + 2 + 0.2(2) = 0 + 2 + 0.4 = 2.4.$$
$$C_3 = \text{mín}\{4, 2.4\} = 2.4.$$

Por lo tanto, si el nivel del inventario baja a cero al entrar al periodo 3 (y se debe producir en ese tiempo), la producción del periodo 3 debe cubrir la demanda de los periodos 3 y 4.

Para determinar C_2 deben considerarse tres casos respecto de la primera vez que el inventario llega a cero después del periodo 2, que puede ocurrir 1) al final del segundo periodo, 2) al final del tercer periodo o 3) al final del cuarto periodo. En la relación recursiva de C_2, estos casos corresponden a 1) $j = 2$, 2) $j = 3$ y 3) $j = 4$, donde los costos correspondientes son $C_2^{(2)}$, $C_2^{(3)}$ y $C_2^{(4)}$, respectivamente. En consecuencia, el costo C_2 es el mínimo entre $C_2^{(2)}$, $C_2^{(3)}$ y $C_2^{(4)}$.

$$C_2^{(2)} = C_3 + 2 = 2.4 + 2 = 4.4.$$
$$C_2^{(3)} = C_4 + 2 + 0.2(3) = 2 + 2 + 0.6 = 4.6.$$
$$C_2^{(4)} = C_5 + 2 + 0.2[3 + 2(2)] = 0 + 2 + 1.4 = 3.4.$$
$$C_2 = \text{mín}\{4.4, 4.6, 3.4\} = 3.4.$$

[5] De todas maneras, el enfoque completo de programación dinámica es útil para resolver *generalizaciones* del modelo (como las funciones de costo de producción y costo de mantener *no lineales*) donde ya no se puede aplicar el algoritmo anterior. (Vea los problemas 18.4-3 y 18.4-4 como ejemplos en los que se utiliza programación dinámica para manejar las generalizaciones del modelo.)

En consecuencia, si la producción se realiza en el periodo 2 (porque el nivel de inventario baja a cero), debe cubrir la demanda de todos los periodos que faltan.

Por último, para encontrar C_1, debe considerarse cuatro casos respecto de la primera vez que el inventario llega a cero, que pueden ocurrir 1) al final del primer periodo, 2) al final del segundo periodo, 3) al final del tercer periodo o 4) al final del cuarto periodo. Estos casos corresponden a $j = 1, 2, 3, 4$ y a los correspondientes costos $C_1^{(1)}$, $C_1^{(2)}$, $C_1^{(3)}$ y $C_1^{(4)}$, respectivamente. El costo C_1 es el mínimo de $C_1^{(1)}$, $C_1^{(2)}$, $C_1^{(3)}$ y $C_1^{(4)}$.

$$C_1^{(1)} = C_2 + 2 = 3.4 + 2 = 5.4.$$
$$C_1^{(2)} = C_3 + 2 + 0.2(2) = 2.4 + 2 + 0.4 = 4.8.$$
$$C_1^{(3)} = C_4 + 2 + 0.2[2 + 2(3)] = 2 + 2 + 1.6 = 5.6.$$
$$C_1^{(4)} = C_5 + 2 + 0.2[2 + 2(3) + 3(2)] = 0 + 2 + 2.8 = 4.8.$$
$$C_1 = \text{mín}\{5.4, 4.8, 5.6, 4.8\} = 4.8.$$

Observe que $C_1^{(2)}$ y $C_1^{(4)}$ empatan como mínimos, lo que da C_1. Esto significa que las políticas correspondientes a $C_1^{(2)}$ y $C_1^{(4)}$ empatan como políticas óptimas. La política $C_1^{(4)}$ establece que se debe producir suficiente en el periodo 1, para cubrir la demanda de los cuatro periodos. La política $C_1^{(2)}$ cubre sólo la demanda hasta el periodo 2. Como esta última política incluye un inventario de cero al final del periodo 2, el resultado C_3 se usa después, es decir, producir suficiente en el periodo 3 para cubrir la demanda de los periodos 3 y 4. Los programas de producción que resultan se resumen a continuación.

Programas de producción óptimos

1. Producir 10 aviones en el periodo 1.
Costo total = \$4.8 millones.

2. Producir 5 aviones en el periodo 1 y 5 en el periodo 3.
Costo total = \$4.8 millones.

Si desea ver un ejemplo más pequeño de la aplicación de este algoritmo, se proporciona uno en la sección Worked Examples en el sitio en internet de este libro.

18.5 MODELOS DE INVENTARIO DETERMINÍSTICOS CON MÚLTIPLES ESCALONES PARA ADMINISTRAR UNA CADENA DE PROVEEDORES

En años recientes, la creciente economía global ha ocasionado un cambio drástico en la administración de inventarios. Hoy, más que nunca, el inventario de muchos fabricantes está disperso por el mundo. Incluso el inventario de un determinado producto puede estar globalmente disperso.

Al principio este inventario se puede guardar en el punto o puntos de manufactura (un *escalón* del sistema de inventario), después en almacenes regionales o nacionales (un segundo escalón), posteriormente en centros de distribución (tercer escalón), y así sucesivamente. De esta forma, cada etapa en la que se retiene el inventario en la progresión a través de un sistema de inventarios con múltiples etapas se llama **escalón** del sistema de inventarios. Tal sistema con múltiples escalones se conoce como **sistema de inventario con escalones múltiples**. En el caso de una corporación integrada que fabrica y vende sus productos hasta el nivel de las tiendas, sus escalones se extienden hasta abajo, hasta el punto del almacén de las tiendas.

Se necesita cierta coordinación entre los inventarios de cualquier producto en los diferentes escalones. Como el inventario en cada escalón (excepto el primero) se reabastece de los escalones más elevados, el nivel que se necesita en los escalones superiores es afectado por el momento en que deben reabastecer los diferentes puntos de escalones inferiores.

El análisis de sistemas de inventarios de escalones múltiples es un reto importante. No obstante, se ha realizado una cantidad considerable de investigación innovadora (con raíces que se pueden rastrear hasta mediados del siglo XX) para desarrollar modelos de inventarios de escalones múltiples manejables. Dada la prominencia creciente de los sistemas de inventarios de escalones múltiples, ésta continuará sin duda como un área de investigación activa.

Recuadro de aplicación

Fundada en 1837, **Deere & Company** es un productor líder a nivel mundial de maquinaria para la agricultura, los bosques y para uso del cliente. La compañía emplea alrededor de 43 000 personas y vende sus productos a través de una red internacional de distribuidores y minoristas.

Por décadas, la división Commercial and Consumer Equipment (C&CE) de Deere enviaba sus inventarios a sus distribuidores, cobraba las ganancias y vivía con la esperanza de que sus distribuidores tuvieran los productos que los clientes demandaban para vendérselos en el momento preciso. Sin embargo, en 2001 la división tenía una razón inventario/ventas anuales de 58% con base en los inventarios propios y de sus distribuidores, por lo que los costos por inventarios escaparon con rapidez de control. Irónicamente, aunque los distribuidores tenían inventarios muy grandes, a menudo no contaban con los productos que los clientes necesitaban.

Los gerentes de la cadena de suministro de la empresa debían reducir los niveles de inventarios y, a la vez, mejorar la disponibilidad del producto y la eficiencia en la entrega. Habían leído acerca de los éxitos de otras compañías en la optimización de inventarios en la revista *Fortune*, por lo que contrataron a una firma líder en consultoría en IO (SmartOps) para cumplir con este reto. Con 300 productos, 2 500 distribuidores en Norteamérica, cinco plantas y sus bodegas correspondientes, siete bodegas en Europa y varios depósitos de consignación de minoristas, la coordinación y optimización de la cadena de suministro de C&CE era, en realidad, un verdadero desafío.

Sin embargo, SmartOps enfrentó estos retos de una manera exitosa mediante la aplicación de técnicas muy modernas en *optimización de inventarios incluidas en su paquete de software para la planeación y optimización de inventarios multietapa* con el fin de establecer objetivos confiables. C&CE utilizó dichos objetivos, junto con los incentivos adecuados para los distribuidores, con el fin de transformar la operación de toda su cadena de suministro. En el proceso, Deere *mejoró de 63 a 92% los embarques de todas sus fábricas*, a la vez que conservó en un 90% los niveles de servicio al cliente. A finales de 2004, la división C&CE había excedido la meta de **100 000 millones de dólares** *en reducción de inventarios*.

Fuente: Troyer, L., J. Smith, S. Marshall, E. Yaniv, S. Tayur, M. Barkman, A. Kaya y Y. Liu: “Improving Asset Management and Order Fulfillment at Deere & Company’s C&CE Division”, en *Interfaces*, **35**(1): 76-87, enero-febrero de 2005. (En el sitio en internet de este libro —www.mhhe.com/hillier— se proporciona una liga hacia este artículo).

Otro concepto clave que ha surgido en la economía global es el de *administración de la cadena de proveedores*. Este concepto conduce a la administración de un sistema de inventarios de escalones múltiples un paso adelante, pues considera lo que debe suceder para incluir un producto en el sistema de inventarios. Sin embargo, con la administración de inventarios, un objetivo primordial es ganar la batalla competitiva contra otras compañías llevando los productos a los clientes lo más pronto posible.

Una **cadena de proveedores** es una red de instalaciones que procura materia prima, la transforma en bienes intermedios y después en productos finales y, por último, entrega los productos a los clientes a través del sistema de distribución que incluye un sistema de inventarios (tal vez de escalones múltiples). Así, se extiende desde la adquisición, manufactura y distribución con una administración eficaz del inventario como elemento clave. Para satisfacer las órdenes con eficiencia, es necesario entender los enlaces e interrelaciones de todos los elementos clave de la cadena de proveedores. Por lo tanto, la administración integrada de ésta se ha convertido en un factor clave del éxito para algunas de las compañías líderes de la actualidad.

Para ayudar en la administración de la cadena de proveedores, los modelos de inventarios con escalones múltiples deben incluir escalones que incorporen la primera parte de la cadena de proveedores así como los escalones para la distribución del producto final. En este contexto, el primer escalón podría ser el inventario de materias primas o componentes que en cierto momento serán usados para fabricar el producto. Un segundo escalón podría ser el inventario de subensambles que se fabrican a partir de las materias primas o componentes en preparación para un ensamblado posterior que los convertirá en el producto final. Este enfoque podría conducir después a los escalones para la distribución del producto terminado, que se inician con el almacenamiento en el punto o puntos de manufactura, después sigue en los almacenes nacionales o regionales, después en los centros de distribución, y así en forma sucesiva.

El objetivo usual de un modelo de inventarios con escalones múltiples es coordinar los inventarios de los diferentes escalones para minimizar el costo total asociado con el sistema de inventarios con escalones múltiples completo. Éste es un objetivo natural de una corporación totalmente integrada que opera todo el sistema. También podría ser un objetivo a considerar cuando ciertos escalones son administrados ya sea por proveedores o clientes de la compañía. La razón es que un concepto clave de la administración de la cadena de proveedores es que una compañía debe tratar de

desarrollar una relación compartida informal con sus proveedores y clientes que les permita maximizar en forma conjunta su ganancia total. Con frecuencia esta sinergia conduce a la concreción de contratos de suministro mutuamente beneficiosos que permite la reducción del costo total de operación de un sistema de inventarios con escalones múltiples administrado en forma conjunta.

El análisis de los modelos de inventarios de escalones múltiples tiende a ser mucho más complicado que aquellos que se enfocan en una sola instalación y que se consideran en el resto del capítulo. Sin embargo, a continuación se presentan dos modelos de este tipo que son relativamente manejables y que ilustran los conceptos relevantes.

Modelo de sistema serial de dos escalones

El sistema de inventarios de escalones múltiples más simple posible consta de sólo dos escalones y una sola instalación en cada uno de ellos. En la figura 18.7 se presenta un sistema de este tipo, donde el inventario en la instalación 1 se usa para reabastecer en forma periódica la instalación 2. Por ejemplo, la instalación 1 puede ser una fábrica que produce cierto producto con corridas de producción ocasionales, y la instalación 2 puede ser el centro de distribución de ese producto. De manera alternativa, esta última puede ser la fábrica que elabora el producto, y la primera es otra instalación donde los componentes necesarios para fabricar ese producto deben ser a su vez fabricados, o recibidos de los proveedores.

Como los productos de las instalaciones 1 y 2 pueden ser diferentes, se hará referencia a ellos como producto 1 y producto 2, respectivamente. Las unidades del producto 1 y del producto 2 se definen de forma que se requiere exactamente una unidad del producto 1 para obtener una unidad del producto 2. Por ejemplo, si el producto 1 consiste colectivamente en los componentes necesarios para fabricar el producto final (producto 2), el conjunto de los componentes que se requiere para fabricar una unidad del producto final se define como una unidad del producto 1.

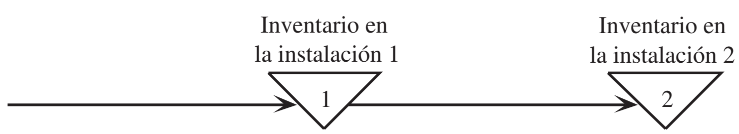
El modelo se basa en los siguientes supuestos.

Supuestos del modelo serial de dos escalones

1. Los supuestos del *modelo EOQ básico* (vea la sección 18.3) se aplican en la instalación 2. Así, existe una tasa de demanda conocida de d unidades por unidad de tiempo, una cantidad por ordenar Q_2 unidades se surte a tiempo para reabastecer el inventario cuando éste llega a un nivel de cero, y no se permiten faltantes planeados.
2. Los costos relevantes en la instalación 2 son un *costo de preparación* de K_2 cada vez que se requiere una orden y un *costo de mantener* de h_2 por unidad por unidad de tiempo.
3. La instalación 1 utiliza su inventario para suministrar un lote de Q_2 unidades a la instalación 2 de manera inmediata cada vez que se recibe una orden.
4. Se ordena una cantidad de Q_1 unidades a tiempo para reabastecer el inventario de la instalación 1 antes de que ocurra un faltante.
5. De manera similar a la instalación 2, los costos relevantes en la instalación 1 son un *costo de preparación* de K_1 cada vez que se requiere una orden y un *costo de mantener* de h_1 por unidad por unidad de tiempo.
6. Las unidades aumentan su valor cuando son recibidas y procesadas en la instalación 2, así $h_1 < h_2$.
7. El objetivo es minimizar la *suma* de los costos variables por unidad de tiempo en las dos instalaciones. (Esto se denotará por C.)

La frase "de inmediato" en el supuesto 3 implica que es esencial que exista un *tiempo de entrega cero* entre el momento en que la instalación 2 coloca una orden por Q_2 unidades y el momento en que la instalación 1 satisface esa orden. En realidad, sería común tener un tiempo de entrega

FIGURA 18.7
Sistema de inventarios serial de dos escalones.



significativo por el tiempo necesario para que la instalación 1 reciba y procese la orden y después transporte el lote a la instalación 2. Sin embargo, como el tiempo de entrega permanece en esencia fijo, para propósitos de evaluación es equivalente suponer que el tiempo de entrega es cero porque la orden será colocada justo a tiempo para que el lote llegue cuando el nivel del inventario caiga hasta cero. Por ejemplo si el tiempo de entrega es una semana, la orden se debería colocar una semana antes de que el nivel del inventario llegue a cero.

Aunque un tiempo de entrega igual a cero y un tiempo de entrega fijo son equivalentes para propósitos de modelado, en este caso se supone un tiempo de entrega de cero específicamente porque esto simplifica la conceptualización de cómo los niveles de inventario en las dos instalaciones varían en forma simultánea a través del tiempo. En la figura 18.8 se muestra esta conceptualización. Debido a que los supuestos del modelo EOQ básico se cumplen en el caso de la instalación 2, los niveles del inventario en ella varían de acuerdo con el patrón final de dientes de sierra que se mostró por primera vez en la figura 18.1. Cada vez que la instalación 2 necesita reabastecer su inventario, la instalación 1 le envía Q_2 unidades del producto 1. Este producto puede ser idéntico al producto 2 (como en el caso de una fábrica que embarca el producto final hacia un centro de distribución). Si no es así (como en el caso de un proveedor que embarca hacia una fábrica los componentes necesarios para elaborar el producto final), la instalación 2 de inmediato utiliza el embarque de Q_2 unidades del producto 1 para fabricar Q_2 unidades del producto 2 (el producto final). Después, el inventario de la instalación 2 se reduce a la tasa de demanda constante de d unidades por unidad de tiempo hasta el siguiente reabastecimiento, el cual ocurre justo cuando el nivel del inventario llega de nuevo a 0.

El patrón de los niveles de inventario a través del tiempo de la instalación 1 es algo más complicado que el de la instalación 2. Es necesario retirar Q_2 unidades del inventario de la instalación 1 para surtir a la instalación 2 cada vez que ésta necesita agregar Q_2 unidades para reabastecer su inventario. Con esto es necesario reabastecer el inventario de la instalación 1 de manera ocasional, así que en forma periódica se coloca una orden por una cantidad de Q_1 unidades. Si se utiliza el

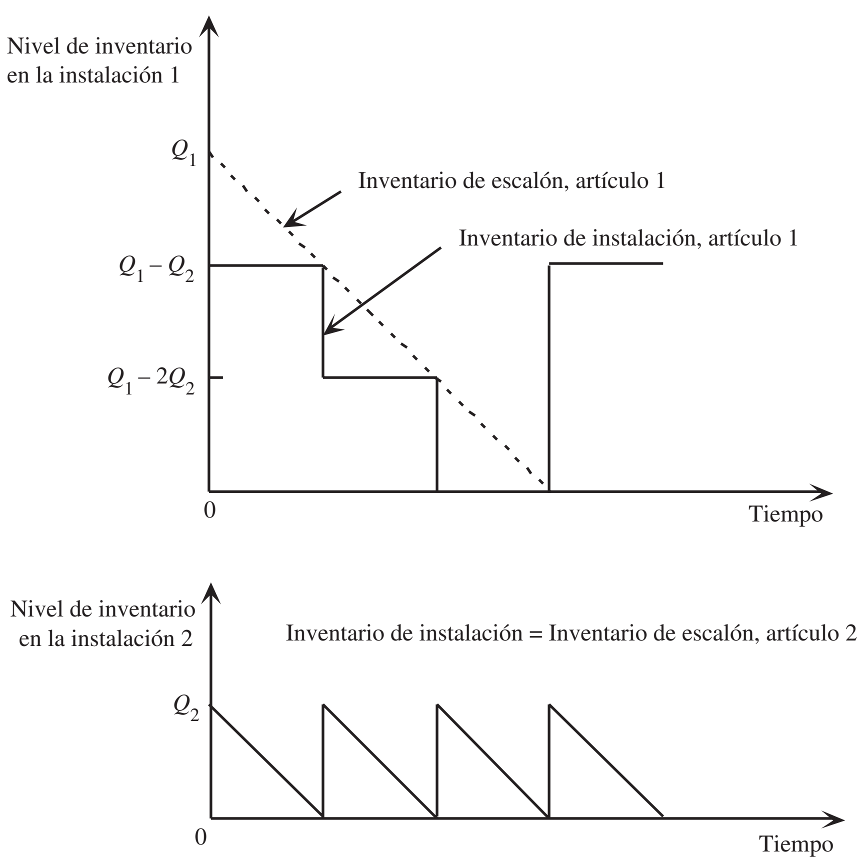


FIGURA 18.8
Niveles de inventario sincronizados en las dos instalaciones cuando $Q_1 = 3Q_2$. El inventario de instalación es aquel que se mantiene físicamente en la instalación, mientras que el inventario de escalón incluye tanto al inventario de instalación como al inventario del mismo artículo que ya está en proceso en la siguiente instalación.

mismo tipo de razonamiento que el que se empleó en la sección anterior (incluso las figuras 18.4 y 18.5), la naturaleza *determinística* de este modelo implica que la instalación 1 debe reabastecer su inventario sólo en el instante en que el nivel de éste es cero y es hora de hacer un retiro de él con el fin de surtir a la instalación 2. El razonamiento implica verificar qué pasaría si la instalación 1 se viese obligada a reabastecer su inventario antes o después de este instante. Si el reabastecimiento fuera después, la instalación 1 podría no surtir a tiempo a la instalación 2 para continuar con la política de inventarios óptima en ese punto, por lo cual sería inaceptable. Si el reabastecimiento se produjera antes de este instante, la instalación 1 incurriría en costo extra por mantener el inventario hasta que sea tiempo de surtir a la instalación 2, así que resulta mejor retrasar el reabastecimiento de la instalación 1 hasta el instante mencionado. Lo anterior conduce al siguiente discernimiento.

> Una política óptima debe tener $Q_1 = nQ_2$, donde n es un entero positivo fijo. Aún más, la instalación 1 debe reabastecer su inventario con un lote de Q_1 unidades *sólo* cuando su nivel de inventario es *cero* y sea tiempo de surtir a la instalación 2 con un lote de Q_2 unidades.

Éste es el tipo de política que se presenta en la figura 18.8, la cual muestra el caso donde $n = 3$. En particular, cada vez que la instalación 1 recibe un lote de Q_1 unidades, en forma simultánea surte a la instalación 2 con un lote de Q_2 unidades, por lo que la cantidad de inventario que queda (llamada *inventario de instalación*) en la instalación 1 es $(Q_1 - Q_2)$ unidades. Después de la provisión de dos lotes más de Q_2 unidades, en la figura 18.8 se muestra que el siguiente ciclo comienza cuando la instalación 1 recibe otro lote de Q_1 unidades al mismo tiempo que necesita surtir a la instalación 2 con otro lote de Q_2 unidades.

La línea punteada de la parte alta de la figura 18.8 muestra otra cantidad llamada *inventario de escalón* de la instalación 1.

> El **inventario de escalón** de un producto particular en cualquier instalación de un sistema de inventarios con escalones múltiples consiste en el inventario del producto que se posee físicamente en la instalación (*inventario de instalación*) *más* el inventario del mismo producto que ya está en escalones subsecuentes del sistema (quizá incorporado a un producto más terminado).

Como el inventario del producto 1 de la instalación 1 se embarca de manera periódica hacia la instalación 2, donde éste se transforma de inmediato en el producto 2, el inventario de escalón de la instalación 1 de la figura 18.8 es la *suma* del inventario que hay en ella y el nivel de inventario en la instalación 2. En el tiempo 0, el inventario de escalón de la instalación 1 es Q_1 porque se conservan $(Q_1 - Q_2)$ unidades y se han embarcado Q_2 unidades a la instalación 2 para reabastecer el inventario de ésta. Como de la instalación 2 se retira el inventario en concordancia con la tasa de demanda constante, el inventario de escalón del producto 1 de la instalación 1 decrece a esta misma tasa constante hasta que se recibe el próximo embarque de Q_1 unidades. Si el inventario de escalón del producto 1 de la instalación 1 se graficara en un periodo más largo que el que se muestra en la figura 18.8, se podría ver el mismo patrón de dientes de sierra para los niveles de inventario como en la figura 18.1.

Pronto se verá que el inventario de escalón tiene un papel fundamental en los sistemas de inventarios de escalones múltiples. La razón es que el patrón de dientes de sierra de los niveles de inventario permite usar un análisis similar al del modelo EOQ básico.

Como el objetivo es minimizar la suma de los costos variables por unidad de tiempo en las dos instalaciones, el enfoque más sencillo (y comúnmente el que más se usa) sería resolver por separado los valores de Q_2 y $Q_1 = nQ_2$ que minimizan el costo variable total por unidad de ambas instalaciones. Desafortunadamente, este enfoque no toma en cuenta (o pasa por alto) las conexiones entre los costos variables en las dos instalaciones. Debido a que el tamaño del lote Q_2 del producto 2 afecta el patrón de los niveles de inventario del producto 1 en la instalación 1, optimizar Q_2 en forma separada sin considerar las consecuencias para el producto 1 no conduce a una solución óptima global.

Para entender mejor este sutil punto, puede resultar instructivo comenzar con la optimización por separado en las dos instalaciones. A continuación se llevará a cabo esta tarea y después se demostrará que ésta es la vía que puede conducir a errores bastante grandes.

La trampa de optimizar las dos instalaciones por separado. Se comenzará por optimizar la instalación 2 por sí misma. Como los supuestos sobre la instalación 2 se ajustan de

manera precisa al modelo EOQ básico, los resultados que se presentaron en la sección 18.3 para este modelo se pueden usar en forma directa. El costo variable total por unidad de tiempo en esta instalación es

$$C_2 = \frac{dK_2}{Q_2} + \frac{h_2 Q_2}{2}.$$

(Esta expresión del costo *variable* total difiere de la del costo total que se dio en la sección 18.3 en el caso del modelo EOQ básico pues se *borra* el costo fijo, dc, donde c es el costo unitario de adquisición del producto.) La fórmula EOQ indica que la cantidad por ordenar óptima para esta instalación por sí misma es

$$Q_2^* = \sqrt{\frac{2dK_2}{h_2}},$$

por lo que el valor de C_2 que resulta con $Q_2 = Q_2^*$ es

$$C_2^* = \sqrt{2dK_2 h_2}.$$

Ahora considere la instalación 1 con una cantidad por ordenar de $Q_1 = nQ_2$. La figura 18.8 indica que el nivel de inventario promedio de la instalación es $(n - 1)Q_2/2$. Por lo tanto, como dicha instalación necesita reabastecer su inventario con Q_1 unidades cada $Q_1/d = nQ_2/d$ unidades de tiempo, el costo variable total por unidad de tiempo en la instalación 1 es

$$C_1 = \frac{dK_1}{nQ_2} + \frac{h_1(n-1)Q_2}{2}.$$

Para determinar la cantidad por ordenar $Q_1 = nQ_2$ que minimiza C_1, dado $Q_2 = Q_2^*$, es necesario resolver para el valor de n que minimiza C_1. Si se pasa por alto el requisito de que n sea un entero, esto se logra al diferenciar C_1 con respecto a n, al establecer la derivada igual a cero (se debe notar que la segunda derivada es positiva para toda n positiva), y al resolver para n, de donde se obtiene

$$n^* = \frac{1}{Q_2^*}\sqrt{\frac{2dK_1}{h_1}} = \sqrt{\frac{K_1 h_2}{K_2 h_1}}.$$

Si n^* es un entero, entonces $Q_1 = n^* Q_2^*$ es la cantidad por ordenar óptima para la instalación 1, dado $Q_2 = Q_2^*$. Si n^* no es un entero, entonces n^* debe redondearse hacia arriba o hacia abajo. La regla para hacer esta operación se presenta a continuación.

Procedimiento de redondeo de *n**

Si $n^* < 1$, se elige $n = 1$.

Si $n^* > 1$, sea $[n^*]$ el entero más grande $\leq n^*$, así que $[n^*] \leq n^* < [n^*] + 1$, y entonces se redondea de la siguiente forma.

$$\text{Si } \frac{n^*}{[n^*]} \leq \frac{[n^*]+1}{n^*}, \text{ elijan } n = [n^*].$$

$$\text{Si } \frac{n^*}{[n^*]} > \frac{[n^*]+1}{n^*}, \text{ elijan } n = [n^*] + 1.$$

La fórmula de n^* indica que su valor depende tanto de K_1/K_2 como de h_2/h_1. Si estas dos cantidades son mucho más grandes que 1, entonces n^* también será mucho más grande que 1. Recuerde que el supuesto 6 del modelo es que $h_1 < h_2$, lo cual implica que h_2/h_1 es mayor que 1, quizá de manera sustancial. La razón por la que usualmente se cumple el supuesto 6 es que por lo general el producto 1 incrementa su valor cuando se convierte en el producto 2 (el producto final) después de que el producto 1 se transfiere a la instalación 2 (la ubicación donde puede satisfacerse la demanda del producto final). Esto significa que el costo de capital incurrido en cada unidad del inventario (por lo general el principal componente del costo de mantener) también se incrementará a medida que las unidades sean trasladadas de la instalación 1 a la instalación 2. En forma similar, si se re-

quiere preparar una corrida de producción para fabricar cada lote en la instalación 1 (por lo que K_1 es grande), mientras que en la instalación 2 sólo se requiere un costo administrativo relativamente pequeño K_2 para colocar cada orden, entonces K_1/K_2 será mucho mayor que 1.

El problema en el análisis anterior surge en el primer paso, cuando se elige la cantidad por ordenar de la instalación 2. En lugar de considerar sólo el costo de la instalación 2 al hacerlo, el costo que resulta en la instalación 1 también debe ser tomado en cuenta. A continuación se presentará un análisis válido que considera de manera simultánea ambas instalaciones al minimizar la suma de los costos de las dos ubicaciones.

Optimización de las dos instalaciones en forma simultánea. Al sumar los costos de cada una de las instalaciones que se obtuvo con anterioridad, el costo variable total por unidad de tiempo en las dos instalaciones es

$$C = C_1 + C_2 = \left(\frac{K_1}{n} + K_2\right)\frac{d}{Q_2} + [(n-1)h_1 + h_2]\frac{Q_2}{2}.$$

El costo de retener de la derecha tiene una interpretación interesante en términos de los costos de mantener en el caso del *inventario de escalón* en las dos instalaciones. En particular, sea

$e_1 = h_1$ = costo unitario de mantener en escalón por unidad de tiempo de la instalación 1,
$e_2 = h_2 - h_1$ = costo unitario de mantener en escalón por unidad de tiempo de la instalación 2.

Entonces, los costos de mantener se pueden expresar como

$$\begin{aligned}[(n-1)h_1 + h_2]\,\frac{Q_2}{2} &= h_1\frac{nQ_2}{2} + (h_2 - h_1)\frac{Q_2}{2}\\ &= e_1\frac{Q_1}{2} + e_2\frac{Q_2}{2},\end{aligned}$$

donde $Q_1/2$ y $Q_2/2$ son los niveles de inventario promedio del *inventario de escalón* de las instalaciones 1 y 2, respectivamente. (Vea la figura 18.8.) La razón por la que $e_2 = h_2 - h_1$ en lugar de $e_2 = h_2$ es que $e_1Q_1/2 = h_1Q_1/2$ ya incluye el costo de mantener de las unidades del producto 1 que están más adelante en el proceso en la instalación 2, por lo que $e_2 = h_2 - h_1$ sólo necesita reflejar el valor agregado que surge de convertir las unidades del producto 1 en unidades del producto 2 en la instalación 2. (Este concepto de utilizar el costo de mantener en escalón basado en el valor agregado en cada instalación tendrá un papel aún más importante en el próximo modelo, donde existen más de dos escalones.)

Si se utilizan estos costos de mantener en escalón, ahora se tiene

$$C = \left(\frac{K_1}{n} + K_2\right)\frac{d}{Q_2} + (ne_1 + e_2)\frac{Q_2}{2}.$$

Al diferenciar con respecto a Q_2, igualar la derivada a cero (mientras se verifica que la segunda derivada es positiva para Q_2 positiva), y resolver para Q_2 se obtiene

$$Q_2^* = \sqrt{\frac{2d\left(\frac{K_1}{n} + K_2\right)}{ne_1 + e_2}}$$

como la cantidad por ordenar óptima (dado n) de la instalación 2. Observe que ésta es idéntica a la fórmula EOQ del modelo EOQ básico donde el costo de preparación total es $K_1/n + K_2$ y el costo total unitario de mantener es $ne_1 + e_2$.

Al insertar esta expresión para Q_2^* en C y realizar algunas simplificaciones algebraicas se obtiene

$$C = \sqrt{2d\left(\frac{K_1}{n} + K_2\right)(ne_1 + e_2)}.$$

Para determinar el valor óptimo de la cantidad por ordenar de la instalación 1, $Q_1 = nQ_2^*$, es necesario encontrar el valor de n que minimiza C. El enfoque usual para hacerlo sería diferenciar C con

respecto a n, igualar esta derivada a cero, y resolver para n. Sin embargo, debido a que la expresión para C implica la obtención de una raíz cuadrada, no es muy conveniente hacerlo de manera directa. Un enfoque más adecuado es eliminar la raíz cuadrada al elevar al cuadrado C y minimizar C^2 en su lugar, puesto que el valor de n que minimiza C^2 también lo hace con C. Por lo tanto, se diferencia C^2 con respecto a n, se iguala esta derivada a cero y se resuelve la ecuación para obtener n. Como la segunda derivada es positiva para n positiva, se obtiene el valor de n que minimiza como

$$n^* = \sqrt{\frac{K_1 e_2}{K_2 e_1}}.$$

Esta ecuación es idéntica a la expresión de n^* que se obtuvo en la subsección anterior excepto que h_1 y h_2 han sido reemplazadas por e_1 y e_2 respectivamente. Cuando n^* no es un entero, el procedimiento para redondearla a un entero es el mismo que se describió en la subsección anterior.

Al obtener n de esta forma es posible calcular Q_2^* con la expresión que se obtuvo con anterioridad y después establecer $Q_1^* = nQ_2^*$.

Ejemplo. Para ilustrar estos resultados, suponga que los parámetros del modelo son

$$K_1 = \$1\,000, \qquad K_2 = \$100, \qquad h_1 = \$2, \qquad h_2 = \$3, \qquad d = 600.$$

En la tabla 18.1 se presentan los valores de Q_2^*, n^*, n (el valor redondeado de n^*), Q_1^* y C^* (el costo variable total por unidad de tiempo que resulta) cuando se resuelve de las dos formas que se describieron en esta sección. Así, la segunda columna da los resultados cuando se utiliza el enfoque impreciso de optimizar las dos instalaciones por separado, mientras que la tercera utiliza el método válido de optimizar las dos instalaciones de manera simultánea.

Observe que este último enfoque produce resultados muy diferentes a los de la optimización separada. La diferencia más grande es que la cantidad por ordenar de la instalación 2 es casi dos veces mayor. Además, el costo variable total C^* es casi 3% más pequeño. Debido a los distintos valores de los parámetros, en ocasiones, el error de la optimización por separado puede conducir a una diferencia porcentual en el costo variable total considerablemente mayor. Por lo tanto, este enfoque proporciona una aproximación muy burda. No existe razón para utilizarlo puesto que la optimización simultánea puede realizarse con facilidad.

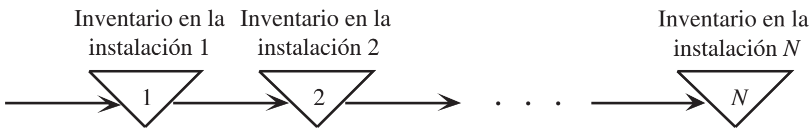
Modelo de un sistema serial con escalones múltiples

A continuación se extenderá el análisis anterior a sistemas seriales con más de dos escalones. En la figura 18.9 se muestra este tipo de sistema, donde la instalación 1 tiene un inventario que se reabastece en forma periódica, que después se usa para surtir el inventario de la instalación 2 de la misma manera periódica, después éste se emplea para surtir la instalación 3, y así sucesivamente hasta la instalación final (instalación N). Algunas de las instalaciones podrían ser centros de procesamiento que manipulan los productos recibidos de la instalación precedente y los transforman en algo más cercano al producto terminado. También se usan para almacenar artículos hasta que éstos están listos para ser trasladados a la siguiente instalación de almacenamiento que está más cercana a los clientes que requieren el producto final. La instalación N lleva a cabo cualquier procesamiento final necesario y también almacena el producto terminado en una ubicación que puede satisfacer de inmediato la demanda continua de ese producto.

TABLA 18.1 Aplicación del modelo serial de dos escalones al ejemplo

Cantidad	Optimización separada de las instalaciones	Optimización simultánea de las instalaciones
Q_2^*	200	379
n^*	$\sqrt{15}$	$\sqrt{5}$
n	4	2
Q_1^*	800	758
C^*	\$1 950	\$1 897

■ **FIGURA 18.9**
Sistema de inventarios serial de escalones múltiples.



Como los productos pueden ser diferentes en las distintas instalaciones en las que son procesados para convertirse en algo más cercano al producto terminado, éstos serán denominados como producto 1 mientras estén en la instalación 1, producto 2 mientras estén en la instalación 2, etc. Las unidades de los distintos productos se definen de manera que se requiera exactamente una unidad del producto en una instalación para obtener una unidad del siguiente producto en la siguiente.

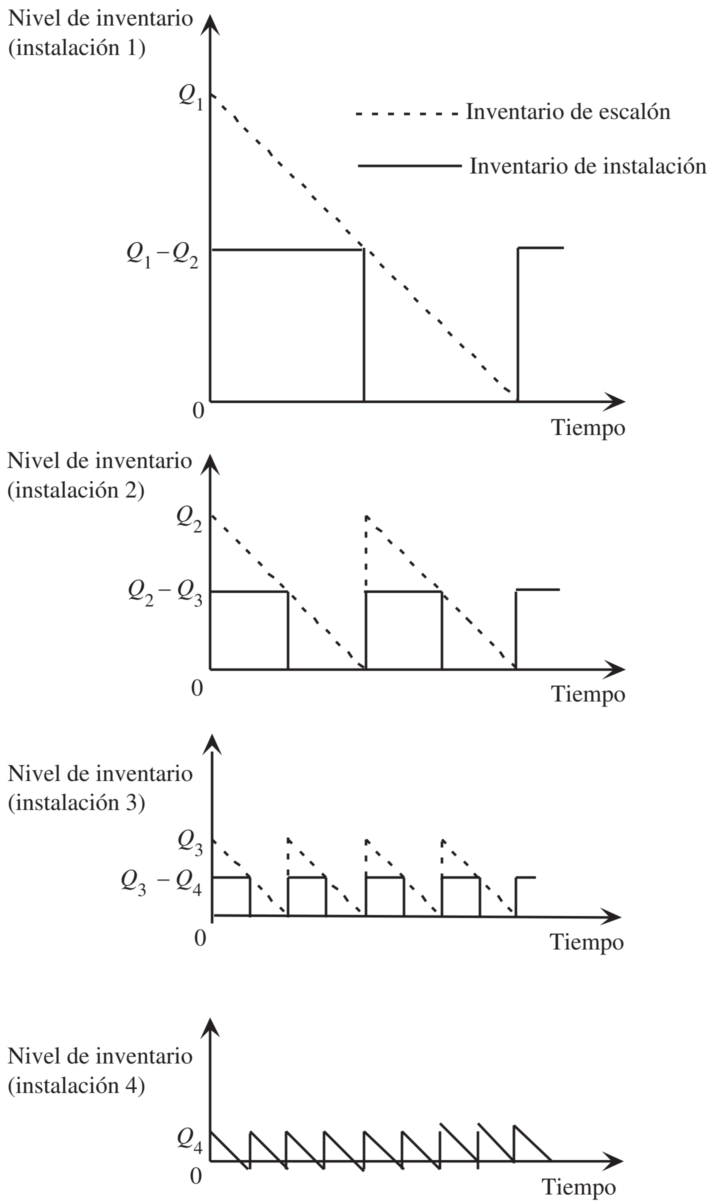
El presente modelo de un sistema de inventarios serial con escalones múltiples es una generalización directa del anterior de un sistema serial de dos escalones, como lo indican los siguientes supuestos del modelo.

Supuestos del modelo serial de escalones múltiples

1. Los supuestos del modelo EOQ básico (vea la sección 18.3) se cumplen en la instalación N. Por lo tanto, existe una demanda constante de d unidades por unidad de tiempo, se coloca una cantidad ordenada de Q_N unidades a tiempo para reabastecer el inventario cuando éste cae hasta cero y no se permiten faltantes planeados.
2. Se coloca una cantidad ordenada de Q_1 unidades a tiempo para reabastecer el inventario en la instalación 1 antes de que se presente un faltante.
3. Cada instalación, excepto la N, utiliza su inventario para reabastecer en forma periódica el inventario de la siguiente instalación. En consecuencia, la instalación i ($i = 1, 2, \ldots, N - 1$) proporciona un lote de Q_{i+1} unidades a la instalación ($i + 1$) de forma inmediata cada vez que se recibe una orden de la instalación ($i + 1$).
4. Los costos relevantes en cada instalación i ($i = 1, 2, \ldots, N$) son un *costo de preparación* de K_i cada vez que se coloca una orden y un costo unitario de *mantener* de h_i por unidad de tiempo.
5. Las unidades incrementan su valor cada vez que se reciben y procesan en la siguiente instalación, por lo cual $h_1 < h_2 < \cdots h_N$.
6. El objetivo es minimizar la *suma* de los costos variables por unidad de tiempo en las N instalaciones. (Esta suma se denotará por C.)

La frase "de inmediato" del supuesto 3 implica que en esencia existe un tiempo de entrega de cero entre el momento en que una instalación coloca una orden y el instante en que la instalación anterior la satisface, aunque un tiempo de entrega positivo que sea fijo no causa ninguna complicación. Con un tiempo de entrega igual a cero, la figura 18.10 es una extensión de la figura 18.8 para mostrar cómo varían los niveles de inventario de manera simultánea en las instalaciones cuando hay cuatro instalaciones en vez de sólo dos. En este caso $Q_i = 2Q_{i+1}$ para $i = 1, 2, 3$, de forma que cada una de las primeras tres instalaciones necesita reabastecer su inventario una vez por cada dos que surte el inventario a la siguiente instalación. En consecuencia, cuando en el tiempo 0 comienza un ciclo completo de reabastecimientos de las cuatro instalaciones, la figura 18.10 muestra una orden de Q_1 unidades que llegan a la instalación 1 cuando el nivel de inventario llega a cero. La mitad de esta orden se usa de inmediato para reabastecer el inventario de la instalación 2. La instalación 2 hace lo mismo con respecto a la instalación 3, y la instalación 3 lo hace con la 4. Por lo tanto, en el tiempo 0, algunas de las unidades que llegan en ese momento a la instalación 1 se transfieren de manera sucesiva hasta la última instalación tan pronto como sea posible. Entonces, la instalación comienza a usar su inventario reabastecido del producto final para satisfacer la demanda de d unidades por unidad de tiempo de ese producto.

Recuerde que el *inventario de escalón* de la instalación 1 está definido como el inventario que se encuentra físicamente en ella (el *inventario de instalación*) más el inventario que ya avanzó en el proceso y se encuentra en escalones posteriores del sistema de inventarios (y quizá ya está incorporado a un producto más terminado). Por lo tanto, como lo indican las líneas punteadas de la figura 18.10, el inventario de escalón de la instalación 1 comienza con Q_1 unidades en el tiempo 0



■ **FIGURA 18.10**
Niveles de inventario sincronizados en las cuatro instalaciones ($N = 4$) cuando $Q_i = 2Q_{i+1}$ ($i = 1, 2, 3$), donde las líneas continuas muestran los niveles de inventario de instalación y las líneas punteadas hacen lo mismo en el inventario de escalón.

y después disminuye a una tasa de d unidades por unidad de tiempo hasta que es tiempo de ordenar otro lote de Q_1, luego de lo cual continúa el patrón de dientes de sierra. El inventario de escalón de las instalaciones 2 y 3 sigue el mismo patrón, pero con ciclos más cortos. Asimismo, coincide con el inventario de instalación de la instalación 4, de forma que aquí, de nuevo, sigue un patrón de dientes de sierra.

Este patrón de dientes de sierra del modelo EOQ básico de la sección 18.3 hace que el análisis sea particularmente directo. Por la misma razón, es conveniente enfocarse en el inventario de escalón en vez de hacerlo en el inventario de instalación de las respectivas instalaciones cuando se analiza el modelo actual. Para hacerlo, es necesario usar los *costos de mantener en escalón*,

$$e_1 = h_1, \quad e_2 = h_2 - h_1, \quad e_3 = h_3 - h_2, \ldots, \quad e_N = h_N - h_{N-1},$$

donde e_i se interpreta como el costo unitario de mantener por unidad de tiempo sobre el *valor agregado* por convertir el producto $(i - 1)$ de la instalación $(i - 1)$ del producto i de la instalación i.

En la figura 18.10 se supone que los ciclos de reabastecimiento de las instalaciones respectivas están sincronizados de manera cuidadosa para que, por ejemplo, el reabastecimiento de la instalación 1 ocurra al mismo tiempo que algunos de los reabastecimientos de las otras instalaciones. Este supuesto tiene sentido puesto que sería oneroso reabastecer el inventario de una instalación antes de que ello fuese necesario. Para evitar tener inventario sobrante al final del ciclo de reabastecimiento en una instalación, también resulta lógico ordenar sólo lo suficiente para suministrar a la siguiente instalación un número entero de veces.

Una política óptima debe tener $Q_i = n_i Q_{i+1}$ ($i = 1, 2, \ldots, N-1$), donde n_i es un entero positivo, en cualquier ciclo de reabastecimiento. (El valor de n_i puede ser diferente en distintos ciclos de reabastecimiento.) Aún más, la instalación i ($i = 1, 2, \ldots, N-1$) debe reabastecer su inventario con un lote de Q_i unidades *sólo* cuando su nivel de inventario es *cero* y es tiempo de surtir a la instalación ($i + 1$) con un lote de Q_{i+1} unidades.

Un problema modificado que es fácil de resolver. Desafortunadamente, resulta muy difícil encontrar una solución óptima para este modelo cuando $N > 2$. Por ejemplo, una solución óptima puede implicar cantidades por ordenar que cambien de un ciclo de reabastecimiento al siguiente en la misma instalación. Por lo tanto, regularmente se hacen dos aproximaciones de simplificación para obtener una solución.

Primera aproximación de simplificación: Suponga que la cantidad que debe ordenar una instalación debe ser la misma en todos los ciclos de reabastecimiento. Así, $Q_i = n_i Q_{i+1}$ ($i = 1, 2, \ldots, N-1$), donde n_i es un entero positivo *fijo*.

Segunda aproximación de simplificación: $n_i = 2^{m_i}$ ($i = 1, 2, \ldots, N-1$), donde m_i es un entero no negativo, de forma que los únicos valores considerados de n_i son 1, 2, 4, 8, …

En efecto, estas aproximaciones de simplificación modifican el problema original pues imponen nuevas restricciones, las cuales reducen el tamaño de la región factible que debe considerarse. Este problema modificado tiene algunas estructuras adicionales (entre ellas el calendario relativamente simple que implica la segunda aproximación de simplificación) que lo hacen mucho más fácil de resolver que el problema original. Además, se ha demostrado que la solución óptima del problema modificado siempre está cerca de ser óptima para el problema original, debido al siguiente resultado clave.

Propiedad de aproximación de 98% de Roundy: El problema modificado *garantiza* obtener una aproximación de al menos 98% del problema original en el sentido siguiente. La cantidad por la que el costo de la solución óptima para el problema modificado excede el costo de la solución óptima para el problema original *nunca* es de más de 2% (y por lo general es mucho menor). En específico, si

C^* = costo variable total por unidad de tiempo de una solución óptima para el problema original,

$\overline{C}$ = costo variable total por unidad de tiempo de una solución óptima para el problema modificado,

entonces

$\overline{C} - C^* \leq 0.02\, C^*$.

Con frecuencia esto se conoce como aproximación de 98% de *Roundy* porque la formulación y prueba de esta propiedad fundamental (que también se cumple para algunos tipos más generales de sistemas de inventarios con escalones múltiples) fue desarrollada por el profesor Robin Roundy, de Cornell University.[6]

Una implicación de las dos aproximaciones de simplificación es que las cantidades que se debe ordenar en el problema modificado deben satisfacer las desigualdades débiles,

$$Q_1 \geq Q_2 \geq \cdots \geq Q_N.$$

[6] R. Roundy, "A 98%-Effective Lot-Sizing Rule for a Muti-Product, Multi-Stage Production/Inventory System", en *Mathematics of Operations Research*, **11**: 699-727, 1986.

El procedimiento para resolver el problema modificado tiene dos fases, donde estas desigualdades tienen un papel clave en la fase 1. En particular, considere la siguiente variación tanto del problema original como del modificado.

> **Una relajación del problema:** Continúe con el supuesto de que la cantidad por ordenar para una instalación debe ser la misma en cada ciclo de reabastecimiento. Sin embargo, reemplace la segunda aproximación de simplificación por el requisito menos restrictivo de que $Q_1 \geq Q_2 \geq \ldots \geq Q_N$. Por lo tanto, la única restricción sobre n_i en la primera aproximación de simplificación es que cada $n_i \geq 1$ ($i = 1, 2, \ldots, N - 1$), sin siquiera requerir que n_i sea un entero. Cuando n_i no es un entero, la falta de sincronización resultante entre las instalaciones se pasa por alto. En su lugar se supone que cada instalación satisface el modelo EOQ básico con el inventario que se reabastece cuando el inventario de escalón llega a cero, sin que importe lo que hagan las otras instalaciones, de manera que todas ellas se pueden optimizar por separado.

Aunque esta relajación no es una representación realista del problema verdadero porque pasa por alto la necesidad de coordinar los reabastecimientos de las instalaciones (y también subestima los costos de mantener verdaderos), proporciona una aproximación que es muy fácil de resolver.

La fase 1 del procedimiento de solución para resolver el problema modificado consiste en resolver la relajación del problema. Después la fase 2 modifica esta solución al restablecer la segunda aproximación de simplificación.

Las desigualdades débiles, $Q_i \geq Q_{i+1}$ ($i = 1, 2, \ldots, N - 1$), permiten la posibilidad de que $Q_i = Q_{i+1}$. (Esto corresponde a tener $m_i = 0$ en la segunda aproximación de simplificación.) Como lo sugiere la figura 18.10, si $Q_i = Q_{i+1}$, siempre que la instalación ($i + 1$) necesite reabastecer su inventario con Q_{i+1} unidades, la instalación i debe ordenar en forma simultánea la misma cantidad de unidades y entonces (después de cualquier procesamiento necesario) transferir de inmediato el lote completo a la instalación ($i + 1$). Por lo tanto, aunque en realidad se cuente con instalaciones separadas, para propósitos de modelación éstas se pueden tratar como una sola instalación combinada que coloca una orden de $Q_i = Q_{i+1}$ unidades con un costo de preparación de $K_i + K_{i+1}$ y un costo de mantener en escalón de $e_i + e_{i+1}$. Esta combinación de instalaciones (para propósitos de modelado) se incorpora en la fase 1 del procedimiento de solución.

A continuación se describe un esquema de las dos fases del procedimiento de solución.

Fase 1 del procedimiento de solución. Recuerde que el supuesto 6 del modelo indica que el objetivo es minimizar C, el costo variable total por unidad de tiempo en todas las instalaciones. Al utilizar los costos de mantener en escalón, el costo variable total por unidad de tiempo en la instalación i es

$$C_i = \frac{dK_i}{Q_i} + \frac{e_i Q_i}{2}, \qquad \text{para } i = 1, 2, \ldots, N,$$

de manera que

$$C = \sum_{i=1}^{N} C_i.$$

(Esta expresión de C_i supone que el inventario de escalón se reabastece cuando su nivel llega a cero, lo cual se cumple para los problemas original y modificado, pero sólo es una aproximación para la relajación del problema porque la falta de coordinación entre las instalaciones al colocar órdenes tiende a generar reabastecimientos prematuros.) Observe que C_i es justo el costo variable total por unidad de tiempo de una sola instalación que satisface el modelo EOQ básico cuando e_i es el costo relevante de mantener por unidad de tiempo en la instalación. Por lo tanto, al resolver el problema relajado, que sólo requiere las instalaciones por separado (cuando se usan costos de mantener en escalón en vez de los costos de mantener en instalación), la fórmula EOQ se debería usar sólo para obtener la cantidad por ordenar de cada instalación. Puede observarse que lo anterior proporciona una primera aproximación razonable de las cantidades por ordenar óptimas cuando se optimizan las instalaciones de manera simultánea en el problema modificado. Por lo tanto, la aplicación de la fórmula EOQ de este modo es el paso clave en la fase 1 del procedimiento de solución. Después, en

la fase 2 se aplica la coordinación necesaria entre las cantidades por ordenar al aplicar la segunda aproximación de simplificación.

Cuando se aplica la fórmula EOQ a las instalaciones respectivas, surge una situación especial cuando $K_i/e_i < K_{i+1}/e_{i+1}$, dado que esta situación conduciría a $Q_i^* < Q_{i+1}^*$, lo cual está prohibido por la relajación del problema. Para satisfacer la relajación, que requiere que $Q_i \geq Q_{i+1}$, lo mejor que se puede hacer es establecer $Q_i = Q_{i+1}$. Como se describió al final de la subsección anterior, esto implica que las dos instalaciones deben combinarse para propósitos de modelado.

Esquema de la fase 1 (solución de la relajación)

1. Si $\frac{K_i}{e_i} < \frac{K_{i+1}}{e_{i+1}}$ para cualquier $i = 1, 2, \ldots, N-1$, las instalaciones i e $i+1$ se deben considerar como una sola instalación combinada (para propósitos de modelado) con un costo de preparación de $K_i + K_{i+1}$ y un costo unitario de mantener en escalón de $e_i + e_{i+1}$ por unidad de tiempo. Después de la combinación, repita este paso cuantas veces sea necesario para cualquier par de instalaciones consecutivas (el cual puede incluir una instalación combinada). Después enumere de nuevo las instalaciones con N como el número total de instalaciones.
2. Establezca

$$Q_i = \sqrt{\frac{2dK_i}{e_i}}, \qquad \text{para } i = 1, 2, \ldots, N.$$

3. Y

$$C_i = \frac{dK_i}{Q_i} + \frac{e_i Q_i}{2}, \qquad \text{para } i = 1, 2, \ldots, N,$$

$$\underline{C} = \sum_{i=1}^{N} C_i.$$

Fase 2 del procedimiento de solución. La fase 2 se utiliza para coordinar las cantidades por ordenar para obtener un calendario cíclico conveniente de reabastecimiento, como el que se ilustra en la figura 18.10. Esto se hace de manera principal al redondear las cantidades por ordenar que se obtuvieron en la fase 1 para ajustarse al patrón prescrito en las aproximaciones de simplificación. Después de determinar de manera tentativa los valores de $n_i = 2^{m_i}$ tales que $Q_i = n_i Q_{i+1}$ de este modo, el paso final es refinar el valor de Q_N para intentar obtener una solución óptima global para el problema modificado.

Este paso final implica expresar cada Q_i en términos de Q_N. En particular, dada cada n_i tal que $Q_i = n_i Q_{i+1}$, sea p_i el producto,

$$p_i = n_i n_{i+1} \ldots n_{N-1}, \qquad \text{para } i = 1, 2, \ldots, N-1,$$

así que

$$Q_i = p_i Q_N, \qquad \text{para } i = 1, 2, \ldots, N-1,$$

donde $p_N = 1$. Por lo tanto, el costo variable total por unidad de tiempo en todas las instalaciones es

$$C = \sum_{i=1}^{N} \left[\frac{dK_i}{p_i Q_N} + \frac{e_i p_i Q_N}{2} \right].$$

Como C incluye sólo la cantidad por ordenar Q_N, esta expresión también se puede interpretar como el costo variable total por unidad de tiempo en una *sola* instalación de inventario que satisface el modelo EOQ básico con un costo de preparación y un costo unitario de mantener de

$$\text{Costo de preparación} = \sum_{i=1}^{N} \frac{dK_i}{p_i}, \qquad \text{Costo unitario de mantener} = \sum_{i=1}^{N} e_i p_i.$$

Por ende, el valor de Q_N que minimiza C está dado por la fórmula EOQ como

$$Q_N^* = \sqrt{\frac{2d\sum_{i=1}^{N}\frac{K_i}{p_i}}{\sum_{i=1}^{N} e_i p_i}}.$$

Debido a que esta expresión requiere conocer las n_i, la fase 2 comienza por la utilización del valor de Q_N que se calculó en la fase 1 como una aproximación de Q_N^*, y después emplea esta Q_N para determinar las n_i (en forma tentativa), antes de usar esta fórmula para calcular Q_N^*.

Esquema de la fase 2 (solución del problema modificado)

1. Establezca Q_N^* como el valor de Q_N que se obtuvo en la fase 1.

2. Para cada $i = N - 1, N - 2, \ldots, 1$, haga lo siguiente. Use el valor de Q_i que obtuvo en la fase 1 para determinar los valores enteros no negativos de m tales que

$$2^m Q_{i+1}^* \le Q_i < 2^{m+1} Q_{i+1}^*.$$

$$\text{Si } \frac{Q_i}{2^m Q_{i+1}^*} \le \frac{2^{m+1} Q_{i+1}^*}{Q_i}, \qquad \text{establezca } n_i = 2^m \text{ y } Q_i^* = n_i Q_{i+1}^*.$$

$$\text{Si } \frac{Q_i}{2^m Q_{i+1}^*} > \frac{2^{m+1} Q_{i+1}^*}{Q_i}, \qquad \text{establezca } n_i = 2^{m+1} \text{ y } Q_i^* = n_i Q_{i+1}^*.$$

3. Use los valores de n_i que obtuvo en el paso 2 y las fórmulas anteriores de p_i y Q_N^* para calcular Q_N^*. Después utilice esta Q_N^* para repetir el paso 2.[7] Si ninguna de las n_i cambia, use $(Q_1^*, Q_2^*, \ldots, Q_N^*)$ como la solución para el problema modificado y calcule el costo correspondiente $\overline{C}$. Si cualquiera de las n_i cambia, repita el paso 2 (inícielo con la Q_N^* actual) y el paso 3 una vez más. Use la solución resultante y calcule $\overline{C}$.

Este procedimiento proporciona una muy buena solución para el problema modificado. Aunque no se garantiza que la solución sea óptima, con frecuencia lo es y, si no, debe estar cerca de serlo. Como el problema modificado es por sí mismo una aproximación del problema original, la obtención de dicha solución para el problema modificado es muy adecuada para todos los propósitos prácticos. La teoría garantiza que esta solución proporcionará una buena aproximación de una solución óptima para el problema original.

Recuerde que la propiedad de aproximación de 98% de Roundy garantiza que el costo de una solución óptima para el problema revisado está, como máximo, 2% alejado de C^*, el costo de la solución óptima desconocida para el problema original. En la práctica, esta diferencia es usualmente mucho menor a 2%. Si la solución que se obtuvo por el procedimiento anterior no es óptima para el problema modificado, el resultado de Roundy aún garantiza que su costo $\overline{C}$ está cuando mucho alejado 6% de C^*. De nuevo hay que destacar que la diferencia usual en la práctica es mucho menor a 6% y con frecuencia también lo es considerablemente menor a 2%.

Sería útil poder verificar qué tan cerca está $\overline{C}$ en cualquier problema particular aunque C^* sea desconocida. La relajación del problema proporciona un modo fácil de hacerlo. Como el problema relajado no requiere la coordinación de los reabastecimientos de inventarios de las instalaciones, el costo calculado para su solución óptima $\underline{C}$ es un límite inferior de C^*. Aún más, por lo general $\underline{C}$ está *extremadamente* cerca de C^*. Por lo tanto, si se verifica qué tan cerca está $\overline{C}$ de $\underline{C}$ se obtiene una estimación conservadora de qué tan cerca debe estar $\overline{C}$ de C^*, como se resume a continuación.

Relaciones de costo: $\underline{C} \le C^* \le \overline{C}$, así $\overline{C} - C^* \le \overline{C} - \underline{C}$, donde
$\underline{C}$ = costo de una solución óptima para el problema *relajado*,
C^* = costo de una solución óptima (desconocida) para el problema *original*,
$\overline{C}$ = costo de la solución del problema *modificado*.

[7] Una posible complicación que podría evitar la repetición del paso 2 es si $Q_{N-1} < Q_N^*$ con este nuevo valor de Q_N^*. Si esto ocurre, se puede detener el procedimiento y usar el valor previo de $(Q_1^*, Q_2^*, \ldots, Q_N^*)$ como la solución para el problema modificado. Esta misma provisión también se aplica en un intento subsecuente de repetir el paso 2.

En el próximo ejemplo típico se verá que, como $\overline{C} = 1.0047\underline{C}$, se sabe que $\overline{C}$ está cuando mucho alejado 0.47% de C^*.

Ejemplo. Considere un sistema serial con cuatro instalaciones que tiene los costos de preparación y costos unitarios de mantener que se muestran en la tabla 18.2.

Cuando se aplica el modelo, el primer paso es convertir los costos unitarios de mantener h_i de cada instalación en el correspondiente costo de mantener en escalón e_i que refleje el valor agregado en cada instalación. Así que

$$e_1 = h_1 = \$0.50, \qquad e_2 = h_2 - h_1 = \$0.05,$$
$$e_3 = h_3 - h_2 = \$3, \qquad e_4 = h_4 - h_3 = \$4.$$

Ahora es posible aplicar el paso 1 de la fase 1 del procedimiento de solución para comparar cada K_1/e_i con K_{i+1}/e_{i+1}.

$$\frac{K_1}{e_1} = 500, \qquad \frac{K_2}{e_2} = 120, \qquad \frac{K_3}{e_3} = 10, \qquad \frac{K_4}{e_4} = 27.5$$

Estas razones decrecen de izquierda a derecha con la excepción de

$$\frac{K_3}{e_3} = 10 < \frac{K_4}{e_4} = 27.5,$$

por lo cual es necesario tratar a las instalaciones 3 y 4 como una sola instalación combinada para propósitos de modelado. Después de combinar sus costos de preparación y sus costos de mantener en escalón, ahora se cuenta con los datos ajustados que se muestran en la tabla 18.3.

Si se utilizan los datos ajustados, en la tabla 18.4 se muestran los resultados de la aplicación del resto del procedimiento de solución a este ejemplo.

TABLA 18.2 Datos del ejemplo de un sistema de inventarios con cuatro escalones

Instalación i	K_i	h_i	d = 4 000
1	$250	$0.50	
2	$6	$0.55	
3	$30	$3.55	
4	$110	$7.55	

TABLA 18.3 Datos ajustados del ejemplo de cuatro escalones después de combinar las instalaciones 3 y 4 para propósitos de modelados

Instalación i	K_i	e_i	d = 4 000
1	$250	$0.50	
2	$6	$0.05	
3(+ 4)	$140	$7	

TABLA 18.4 Resultados de la aplicación del procedimiento de solución al ejemplo de cuatro escalones

	Solución del problema relajado		Solución inicial del problema		Solución final del problema modificado	
Instalación i	Q_i	C_i	Q_i^*	C_i	Q_i^*	C_i
1	2 000	$1 000	1 600	$1 025	1 700	$1 013
2	980	$49	800	$50	850	$49
3(+ 4)	400	$2 800	400	$2 800	425	$2 805
		$\underline{C}$ = $3 849		C = $3 875		$\overline{C}$ = $3 867

La segunda y tercera columnas presentan los cálculos directos de los pasos 2 y 3 de la fase 1. En el paso 1 de la fase 2, $Q_3 = 400$ en la segunda columna se traslada a $Q_3^* = 400$ en la cuarta columna. En el paso 2, se encuentra que

$$2^1 Q_3^* < Q_2 < 2^2 Q_3^*$$

dado que

$$2(400) = 800 < 980 < 4(400) = 1\,600.$$

Debido a que

$$\frac{Q_2}{2^1 Q_3^*} = \frac{980}{800} < \frac{1\,600}{980} = \frac{2^2 Q_3^*}{Q_2},$$

se establece $n_2 = 2^1 = 2$ y $Q_2^* = n_2 Q_3^* = 800$. En forma similar, se establece $n_1 = 2^1 = 2$ y $Q_1^* = n_1 Q_2^* = 1\,600$, puesto que

$$2(800) = 1\,600 < 2\,000 < 4(800) = 3\,200 \quad \text{y} \quad \frac{2\,000}{1\,600} < \frac{3\,200}{2\,000}.$$

Después de calcular la correspondiente C_i, la cuarta y quinta columnas de la tabla resumen estos resultados de la aplicación de los pasos 1 y 2 de la fase 2.

Las últimas dos columnas de la tabla resumen los resultados de completar el procedimiento de solución mediante la aplicación del paso 3 de la fase 2. Como $p_1 = n_1 n_2 = 4$ y $p_2 = n_2 = 2$, la fórmula de Q_N^* produce $Q_3^* = 425$ como el valor de Q_3 que es parte de la solución óptima global para el problema modificado. Si se repite el paso 2 con esta nueva Q_3^*, de nuevo se obtiene $n_2 = 2$ y $n_1 = 2$, por lo que $Q_2^* = n_2 Q_3^* = 850$ y $Q_1^* = n_1 Q_2^* = 1\,700$. Como n_2 y n_1 no cambian desde la primera vez a través del paso 2, ahora se tiene la solución deseada para el problema modificado, de manera que C_i se calcula en concordancia. (En realidad, esta solución es óptima para el problema modificado.)

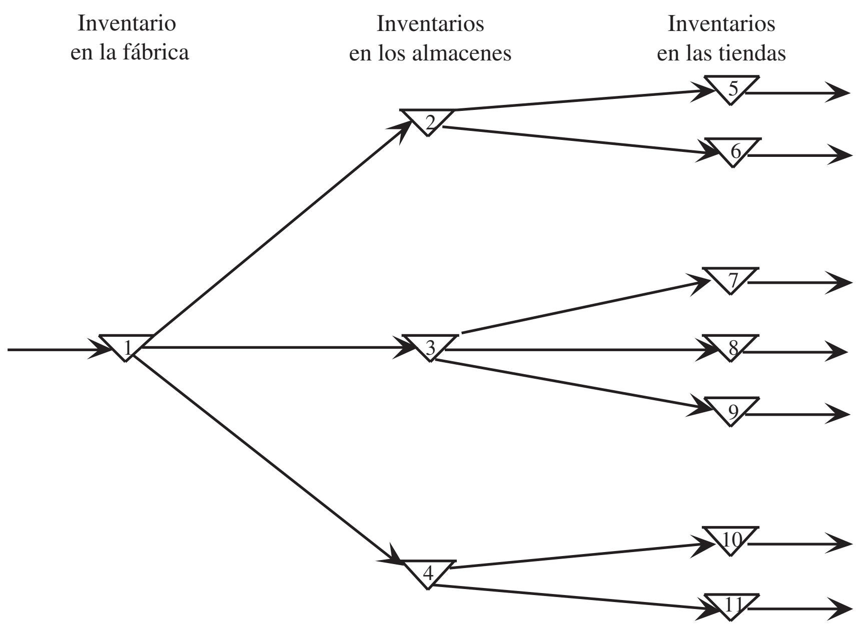
Tenga en mente que las instalaciones originales 3 y 4 se han combinado sólo para propósitos de modelado y por supuesto continuarán como instalaciones físicamente separadas. Por lo tanto, la conclusión en la sexta columna de la tabla es que $Q_3^* = 425$ significa, en realidad, que tanto la instalación 3 como la 4 tendrán una cantidad por ordenar de 425. Tan pronto como la instalación 3 recibe y procesa cada orden, transfiere de inmediato el lote completo a la instalación 4.

La parte baja de la tercera, quinta y séptima columnas de la tabla muestran el costo variable total por unidad de tiempo de las soluciones correspondientes. El costo de $\overline{C}$ en la quinta columna está 0.68% por encima de $\underline{C}$ en la tercera columna, mientras que $\overline{C}$ en la séptima columna está sólo 0.47% por encima de $\underline{C}$. Como $\underline{C}$ es el límite inferior de C^*, el costo de la solución óptima (desconocida) del problema original, esto significa que al detener el procedimiento después del paso 2 de la fase 2 se obtuvo una solución que está cuando mucho 0.68% alejada de C^*, mientras que el refinamiento que implica realizar el paso 3 de la fase 2 mejoró la solución hasta un alejamiento máximo de 0.47% de C^*.

Extensiones de estos modelos

Los dos modelos que se presentaron en esta sección son para sistemas de inventarios seriales. Como se muestra en la figura 18.9, esto restringe cada instalación (después de la primera) a tener sólo un *antecesor inmediato* que reabastezca su inventario. En el mismo sentido, cada instalación (antes de la última) reabastece el inventario de sólo un *sucesor inmediato*.

Muchos sistemas de inventarios con escalones múltiples son más complicados que este ejemplo. Una instalación puede tener *muchos sucesores inmediatos*, como cuando una fábrica surte a muchos almacenes o cuando un almacén abastece a muchas tiendas. Tal sistema de inventarios se llama **sistema de distribución**. En la figura 18.11 se muestra un sistema de inventarios de distribución clásico, el cual prepara una corrida de producción rápida cada vez que necesita reabastecer su inventario del producto. Este inventario se utiliza para surtir a varios almacenes en diferentes regiones, pues puede reabastecer sus inventarios del producto cada vez que se necesita. A su vez,

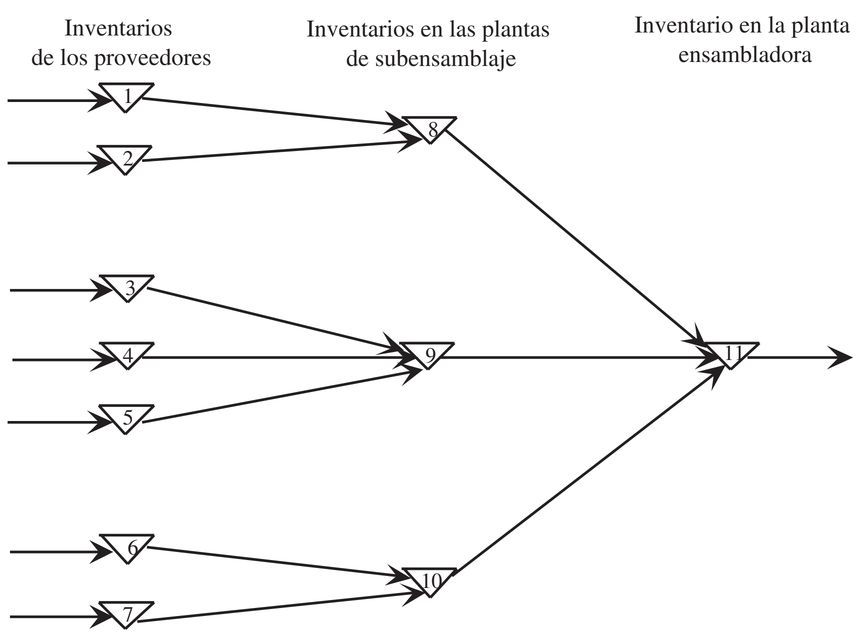


■ **FIGURA 18.11**
Sistema de inventarios de distribución típico.

cada uno de estos almacenes suministra el producto a algunas tiendas dentro de su región, esto es, reabastece sus inventarios cuando es necesario. Si cada tienda tiene una tasa de demanda constante y conocida (al menos de manera aproximada) del producto, es posible formular una extensión del modelo serial de escalones múltiples para este sistema de inventarios de distribución. (No se continuará con este análisis.)

Otra generalización común de un sistema de inventarios con escalones múltiples surge cuando algunas instalaciones tienen *muchos antecesores inmediatos*, como cuando una planta de subensamblado recibe sus componentes de múltiples proveedores o cuando una fábrica recibe sus subensambles de muchas plantas de subensamble. Un sistema de inventarios de este tipo se llama **sistema de ensamblado**. En la figura 18.12 se muestra un sistema de inventarios de ensamblado típico. En este caso, se ensambla un producto particular en una planta de ensamblado, a partir de

■ **FIGURA 18.12**
Sistema de inventarios de ensamblado normal.



inventarios de subensambles que se mantienen en ella para procesar el producto. Cada uno de estos inventarios de subensambles debe ser reabastecido cuando es necesario por una planta que produce ese subensamble, a partir de inventarios de componentes que se mantienen en ella para producir el subensamble. A su vez, cada uno de estos inventarios de un componente debe ser reabastecido cuando le es necesario a un proveedor que produce en forma periódica este componente para reabastecer su propio inventario. Bajo los supuestos apropiados, se puede formular otra extensión del modelo serial de escalones múltiples para este sistema de inventarios de ensamblaje.

Algunos sistemas de inventarios con escalones múltiples pueden incluir tanto instalaciones que tienen sucesores inmediatos múltiples, como antecesores inmediatos múltiples. (Incluso algunas instalaciones pueden caer en ambas categorías.) Algunos de los grandes retos de la administración de cadenas de proveedores se presentan al tratar con estos sistemas de inventarios con escalones múltiples mixtos. Un desafío particular surge cuando organizaciones distintas (por ejemplo, los proveedores, fabricantes y vendedores) controlan partes diferentes de un sistema de inventarios de escalones múltiples, ya sea un sistema mixto, de distribución o de ensamblado. En este caso, un principio clave para administrar una cadena de proveedores es que las organizaciones deben trabajar juntas, incluso a través del desarrollo de contratos de suministro con beneficios mutuos, para optimizar la operación global del sistema de inventarios de escalones múltiples.

Aunque el análisis de los sistemas de distribución y de ensamblado presenta algunas complicaciones adicionales, el enfoque que se presenta aquí del modelo serial de escalones múltiples (incluso la propiedad de aproximación de 98% de Roundy) también se puede extender a estos otros tipos de sistemas de escalones múltiples. En la referencia seleccionada 6 se presentan los detalles (también vea la referencia seleccionada 1 para obtener información adicional acerca de este tipo de sistemas de inventarios, así como mayores detalles sobre los modelos de sistemas seriales.)

Otra forma de extender el modelo serial de escalones múltiples es permitir que la demanda por el producto de la instalación *N* ocurra de manera *aleatoria* en lugar de hacerlo con una tasa de demanda constante. Ésta es un área en la que, en la actualidad, se enfoca la investigación.[8]

18.6 MODELO ESTOCÁSTICO CON REVISIÓN CONTINUA

A continuación se estudiarán los modelos de inventarios *estocásticos*, que están diseñados para analizar sistemas de inventarios donde existe una gran incertidumbre sobre las demandas futuras. En esta sección se considerará un sistema de inventario con *revisión continua*. El nivel del inventario se supervisa en forma continua, por lo que una orden se coloca en cuanto el nivel de inventario llega al punto de reorden.

El método tradicional para implantar un sistema de inventarios de *revisión continua* consistía en usar un **sistema de dos contenedores**. Todas las unidades de cierto producto se colocaban en dos contenedores. La capacidad de uno era igual al punto de reorden. Las unidades se extraían primero del otro contenedor. Entonces, cuando este segundo contenedor se vaciaba, se activaba la señal para colocar una orden. Durante el tiempo de entrega hasta que se recibía la orden, las unidades se extraían del primer contenedor.

En años más recientes, los sistemas de dos contenedores han sido sustituidos por **sistemas de inventarios computarizados**. Se hace un registro electrónico de cada adición al inventario y cada venta que ocasiona una salida, y la computadora siempre tiene el nivel actual del inventario. (Por ejemplo, los dispositivos de lectura de códigos de las cajas registradoras de las tiendas detallan por un lado las compras y por el otro registran la venta de productos para ajustar los niveles de inventario actuales.) La computadora envía una orden en cuanto el nivel de inventario llega al punto de reorden. Se dispone de varios paquetes de software excelentes para implantar sistemas de inventarios de este tipo.

Debido al extenso uso de las computadoras para la administración de inventarios moderna, cada vez se utilizan más los sistemas de revisión continua de productos suficientemente importantes que garantizan una política de inventarios formal.

---

[8] Por ejemplo, consulte H. K. Shang y L.-S. Song, "Newsvendor Bounds and Heuristic for Optimal Policies in Serial Supply Chains", en *Management Science*, **49**(5): 618-638, mayo de 2003. También consulte X. Chao y S. X. Zhou, "Probabilistic Solution and Bounds for Serial Inventory Systems with Discounted and Average Costs", en *Naval Research Logistics*, **54**(6): 623-631, septiembre de 2007.

Es común que un sistema de inventarios de revisión continua de un producto específico se base en dos números críticos:

R = punto de reorden.
Q = cantidad por ordenar.

En el caso de un fabricante que administra su inventario de productos terminados, la orden será para llevar a cabo una *corrida de producción* de tamaño Q. En el de un distribuidor o un comerciante (o un fabricante que reabastece su materia prima con un proveedor), la orden será una *orden de compra* de Q unidades de productos.

Una política de inventarios basada en estos dos números críticos es sencilla.

Política de inventarios: siempre que el nivel de inventario de un producto llegue a R unidades, se coloca una orden de Q unidades para reabastecerlo.

Con frecuencia, esta política se llama *política de punto de reorden*, *cantidad por ordenar*, o **política (R, Q)**. [En consecuencia, el modelo completo se conoce como modelo (R, Q). En ocasiones se usan otras variaciones de estos nombres, como política (Q, R), modelo (Q, R), etcétera.]

Después de resumir los supuestos del modelo, se describe cómo se pueden determinar R y Q.

Supuestos del modelo

1. Cada aplicación se refiere a un solo producto.
2. El nivel de inventario está bajo *revisión continua*, por lo que su valor actual se conoce.
3. Debe usarse una política (R, Q), por lo cual las únicas decisiones que deben tomarse son las selecciones de R y Q.
4. Existe un *tiempo de entrega* entre la colocación de una orden y la recepción de la cantidad ordenada. Este tiempo de entrega puede ser fijo o variable.
5. La *demanda* para retirar unidades del inventario y venderlas (o usarlas de otro modo) durante este tiempo de entrega es incierta. Sin embargo, se conoce (o se puede estimar) la distribución de probabilidad de la demanda.
6. Si ocurren faltantes antes de recibir la orden, el exceso de demanda queda *pendiente*, de manera que estos faltantes se satisfacen cuando llega la orden.
7. Se incurre en *costo de preparación* (denotado por K) cada vez que se coloca una orden.
8. Excepto por este costo fijo, el costo de la orden es proporcional a la cantidad Q.
9. Se incurre en un costo de mantener (denotado por h) por cada unidad en inventario por unidad de tiempo.
10. Cuando ocurren faltantes, se incurre en cierto costo por faltantes (denotado por p) por cada unidad que falta por unidad de tiempo hasta que se satisface la demanda pendiente.

Este modelo tiene una relación estrecha con el *modelo EOQ con faltantes planeados* que se presentó en la sección 18.3. En realidad, todos estos supuestos son congruentes con ese modelo, con la única excepción clave del supuesto 5. En lugar de tener demanda incierta, ese modelo supone una *demanda conocida* con una tasa fija.

Debido a la relación tan cercana entre estos dos modelos, sus resultados deben ser similares. La diferencia principal es que, en razón de la demanda incierta del modelo actual, debe agregarse un inventario de seguridad cuando se establece el punto de reorden para proporcionar un colchón de seguridad si se presenta una demanda por encima de la promedio durante el tiempo de entrega. De otra manera, las compensaciones entre los diferentes factores de costo son en esencia los mismos, por lo cual las cantidades por ordenar en los dos modelos deben ser similares.

Elección de la cantidad de la orden *Q*

El enfoque más directo para elegir Q en el modelo actual es sencillamente usar la fórmula que se explicó en la sección 18.3 para el modelo EOQ con faltantes planeados. Esta fórmula es

$$Q = \sqrt{\frac{2dK}{h}}\sqrt{\frac{p+h}{p}},$$

donde d es ahora la demanda *promedio* por unidad de tiempo, y K, h y p están definidos en los respectivos supuestos 7, 9 y 10.

Esta Q será sólo una aproximación de la cantidad óptima que se debe ordenar según el modelo actual. Sin embargo, no se dispone de una fórmula para determinar el valor exacto de esa cantidad, por lo que se necesita una aproximación. Por fortuna, la que se logra con este método es bastante buena.[9]

Elección del punto de reorden *R*

Un enfoque común para elegir el punto de reorden R se basa en el nivel deseado de servicio al cliente que tenga la administración. Por lo tanto, el punto inicial es obtener una decisión administrativa con respecto al nivel de servicio deseado. (El problema 18.6-3 analiza los factores involucrados en esta decisión administrativa.)

El nivel de servicio se puede definir de varias maneras en este contexto, como se describe a continuación.

Medidas alternativas del nivel de servicio

1. Probabilidad de que ocurra un faltante entre la colocación de la orden y la recepción del pedido.
2. Número promedio de faltantes por año.
3. Porcentaje promedio de la demanda anual que se satisface de inmediato (sin faltantes).
4. Retraso promedio para satisfacer las órdenes pendientes cuando ocurre un faltante.
5. Retraso promedio global para satisfacer las órdenes (donde el retraso sin faltantes es 0).

Las medidas 1 y 2 tienen una relación estrecha. Por ejemplo, suponga que la cantidad por ordenar Q se ha establecido en 10% de la demanda anual, de modo que se coloca un promedio de 10 órdenes por año. Si la probabilidad de *que ocurra* una faltante durante el tiempo de entrega hasta que se recibe la orden es 0.2, entonces el número promedio de faltantes por año sería $10(0.2) = 2$.

Las medidas 2 y 3 también se relacionan. Por ejemplo, suponga que ocurre un promedio de 2 faltantes por año y que el lapso promedio del faltante es de 9 días. Como $2(9) = 18$ días de faltantes por año son en esencia 5% del año, el porcentaje promedio de la demanda anual que se puede satisfacer de inmediato sería 95%.

Además, las medidas 3, 4 y 5 también se relacionan. Por ejemplo, suponga que el porcentaje promedio de la demanda anual que se puede satisfacer de inmediato es 95% y que el retraso promedio para cumplir con las órdenes pendientes cuando ocurren faltantes es de 5 días. Como sólo 5% de los clientes sufren este retraso, el retraso promedio global para satisfacer las órdenes sería $0.05(5) = 0.25$ días por orden.

Debe tomarse una decisión administrativa sobre el valor deseado de al menos una de estas medidas del nivel de servicio. Después de seleccionar una de ellas para estudiarla, es útil explorar las implicaciones de diferentes valores alternativos de esta medida en algunas de las otras medidas, antes de elegir la mejor opción.

La medida 1 tal vez es la más conveniente para usarla como medida principal, por lo que se examinará este caso. Se denotará el nivel de servicio deseado, de acuerdo con esta medida, por L; entonces

$L =$ probabilidad deseada por la administración de que no ocurran faltantes en el lapso entre colocar una orden y recibirla.

Si se usa la medida 1 se tiene que trabajar con la distribución de probabilidad estimada de la siguiente variable aleatoria.

$D =$ demanda durante el tiempo de entrega para satisfacer una orden.

[9] Si desea más información acerca de la calidad de esta aproximación, vea S. Axsäter, “Using the Deterministic EOQ Formula in Stochastic Inventory Control”, en *Management Science*, **42:** 830-834, 1996. Vea también Y.-S. Zheng, “On Properties of Stochastic Systems”, en *Management Science*, **38:** 87-103, 1992.

Por ejemplo, con una distribución uniforme, la fórmula para elegir el punto de reorden R es sencilla.

Si la distribución de probabilidad de D es una *distribución uniforme* en el intervalo de a a b, establezca

$$R = a + L(b - a),$$

porque entonces

$$P(D \leq R) = L.$$

Como la media de la distribución es

$$E(D) = \frac{a + b}{2},$$

la cantidad de **inventario de seguridad** (el nivel de inventario *esperado justo* antes de que la cantidad ordenada se reciba) que proporciona el punto de reorden R es

$$\text{Inventario de seguridad} = R - E(D) = a + L(b - a) - \frac{a + b}{2}$$
$$= \left(L - \frac{1}{2}\right)(b - a).$$

Cuando la distribución de la demanda es diferente a una distribución uniforme, el procedimiento para encontrar R es similar.

Procedimiento general para elegir *R* con la medida 1 de nivel de servicio

1. Seleccionar L.
2. Despejar R tal que

$$P(D \leq R) = L.$$

Por ejemplo, suponga que D tiene distribución normal con media μ y varianza σ^2, como se muestra en la figura 18.13. Dado el valor de L, se puede usar la tabla de distribución normal que se proporciona en el apéndice 5 para determinar el valor de R. En particular, sólo se necesita encontrar el valor de K_{1-L} en esta tabla y después sustituirlo en la siguiente fórmula para calcular R.

$$R = \mu + K_{1-L}\sigma.$$

La cantidad de inventario de seguridad que se obtiene es

$$\text{Inventario de seguridad} = R - \mu = K_{1-L}\sigma.$$

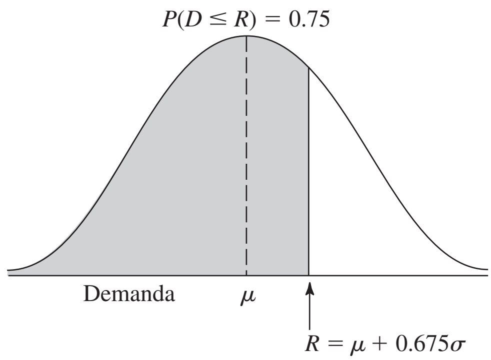
Como ejemplo, si $L = 0.75$, entonces $K_{1-L} = 0.675$, y

$$R = \mu + 0.675\sigma,$$

como se muestra en la figura 18.13. De aquí se obtiene

$$\text{Inventario de seguridad} = 0.675\sigma.$$

En el OR Courseware también se incluye una plantilla de Excel que calcula tanto la cantidad por ordenar como el punto de reorden R. Debe introducirse la demanda promedio por unidad de tiempo (d), los costos (K, h y p) y el nivel de servicio basado en la medida 1. También se indica si la distribución de probabilidad de la demanda durante el tiempo de entrega es una distribución uniforme o normal. Para una distribución uniforme, se especifica el intervalo en el que se extiende la distribución con el límite inferior y el límite superior de este intervalo. En el caso de la distribución normal, se proporciona la media μ y la desviación estándar σ de la distribución. Después de introducir esta información, la plantilla calcula de inmediato Q y R y despliega los resultados en el lado derecho.



■ **FIGURA 18.13**
Cálculo del punto de reorden R según el modelo estocástico con revisión continua, cuando $L = 0.75$ y la distribución de probabilidad de la demanda en el tiempo de entrega es una normal con media μ y desviación estándar σ.

Ejemplo

Considere de nuevo el ejemplo 1 (fabricación de bocinas para televisores) de la sección 18.1. Recuerde que el costo de preparación para producir las bocinas es $K = \$12\,000$, el costo unitario de mantener es $h = \$0.30$ mensual por bocina y el costo unitario por faltantes es $p = \$1.10$ por bocina por mes.

La tasa de demanda fija original era de 8 000 bocinas por mes para ensamblarse en los televisores fabricados en la línea de producción a esta tasa fija. Sin embargo, las ventas han sido muy variables, por lo que el nivel del inventario de televisores terminados fluctúa de manera importante. Para reducir los costos de mantener el inventario de productos terminados, la administración ha decidido ajustar la tasa de producción diaria para que se ajuste mejor a las órdenes que se reciben.

En consecuencia, la demanda actual de bocinas es variable. Existe un *tiempo de entrega* de 1 mes entre ordenar una corrida de producción de bocinas y tenerlas listas para el ensamblado. La demanda de bocinas durante este tiempo de entrega es una variable aleatoria D que tiene distribución normal con media de 8 000 y desviación estándar de 2 000. Para minimizar el riesgo de interrumpir la línea de producción de televisores, la administración ha decidido que el inventario de seguridad de bocinas debe ser suficiente para evitar faltantes 95% del tiempo durante este periodo de entrega.

Para aplicar el modelo, la cantidad que es necesario ordenar en cada corrida de producción debe ser

$$Q = \sqrt{\frac{2dK}{h}}\sqrt{\frac{p+h}{p}} = \sqrt{\frac{2(8\,000)(12\,000)}{0.30}}\sqrt{\frac{1.1+0.3}{1.1}} = 28\,540.$$

Ésta es la misma cantidad por ordenar que se determinó con el modelo EOQ con faltantes planeados en la sección 18.3 en el caso de la versión anterior de este ejemplo, donde se tenía una tasa de demanda *constante* (en lugar de promedio) de 8 000 bocinas por mes con faltantes planeados permitidos. Sin embargo, la diferencia clave es que ahora debe proporcionarse un inventario de seguridad para contrarrestar la demanda variable. La administración ha elegido un nivel de servicio de $L = 0.95$, y de la tabla de la normal en el apéndice 5 se obtiene $K_{1-L} = 1.645$. Por lo tanto, el punto de reorden debe ser

$$R = \mu + K_{1-L}\sigma = 8\,000 + 1.645(2\,000) = 11\,290.$$

La cantidad que resulta como inventario de seguridad es

$$\text{Inventario de seguridad} = R - \mu = 3\,290.$$

En la sección Worked Examples del sitio en internet de este libro se proporciona **otro ejemplo** de la aplicación de este modelo cuando se tiene la posibilidad de utilizar dos opciones de embarque con diferentes distribuciones para el tiempo de entrega y debe identificarse la opción menos costosa.

18.7 MODELO ESTOCÁSTICO DE UN SOLO PERIODO PARA PRODUCTOS PERECEDEROS

Cuando se elige el modelo de inventarios que se debe usar para un producto dado, debe distinguirse entre dos tipos de productos. Uno de ellos es un **producto estable**, que conservará sus ventas en forma indefinida, por lo que no hay una fecha establecida para agotar el inventario. Éste es el tipo de producto que se considera en las secciones anteriores (igual que en la siguiente). El otro tipo, por el contrario, es un **producto perecedero**, que se puede tener en inventario sólo un periodo limitado antes de que no se pueda vender. Éste es el tipo de producto para el que se diseñó el modelo de un solo periodo (y sus variaciones) que se presentó en esta sección. En particular, el único periodo del modelo es el periodo muy limitado antes de que no sea posible vender el producto.

Un ejemplo de un producto perecedero es el periódico del día que se vende en los puestos. El ejemplar de un día determinado se puede tener en inventario sólo un día antes de que caiga en la obsolescencia y deba sustituirse por el del día siguiente. Cuando la demanda del periódico es una variable aleatoria (como se supone en esta sección), el dueño del puesto debe elegir una cantidad por ordenar diaria que proporcione un equilibrio entre el costo potencial de ordenar más de lo necesario (el gasto de desperdicio por ordenar más periódicos de los que se pueden vender) y el costo potencial de ordenar menos (la ganancia perdida por ordenar menos periódicos de los que se pueden vender). El modelo de esta sección permite obtener la cantidad diaria por ordenar que maximice la ganancia esperada.

Como el problema general que se analiza se ajusta a este ejemplo, por tradición este problema ha recibido el nombre de **problema del voceador**. Sin embargo, siempre se ha reconocido que el modelo se aplica a otros productos perecederos. En realidad, la mayoría de las aplicaciones se han realizado para productos perecederos distintos de los periódicos, incluso los ejemplos de este tipo de productos que se enumeran a continuación.

Algunos tipos de productos perecederos

Al revisar la lista siguiente de los diferentes tipos de productos perecederos, puede observarse la similitud que hay entre administrar el inventario de ellos y manejar el número de diarios en el puesto de venta, debido a que estos productos tampoco se pueden vender después de cierto periodo. En lo que pueden diferir es que la longitud de este periodo puede ser una semana, un mes o incluso varios meses en lugar de sólo un día.

1. Publicaciones periódicas, como revistas y periódicos.
2. Flores que vende una florería.
3. La elaboración de comida fresca preparada en un restaurante.
4. Frutas y vegetales que se venden en un supermercado.
5. Árboles de Navidad.
6. Ropa de temporada, como abrigos de invierno, donde las piezas que quedan al final de la temporada deben venderse con un gran descuento para tener espacio para la siguiente estación.
7. Tarjetas de felicitación de temporada.
8. Bienes de moda que pronto estarán fuera de uso.
9. Automóviles nuevos al final del año que corresponde al modelo.
10. Cualquier producto de rápida obsolescencia.
11. Refacciones vitales que deben producirse durante la última corrida de producción de cierto modelo de un producto (como un avión) para usarse cuando se requiera durante la vida útil del modelo.
12. Las reservaciones en una línea aérea para un vuelo específico, puesto que los asientos disponibles en el vuelo representan el inventario de un producto perecedero (no se pueden vender una vez que despega el avión).

Este último tipo es particularmente interesante porque las líneas aéreas principales (y otras compañías relacionadas con el transporte de pasajeros) hacen amplio uso de la IO para analizar la forma de maximizar su ganancia cuando tienen este tipo especial de inventario. Esta rama especial de la teoría de inventarios (conocida con el nombre de *administración de la ganancia*) es el tema principal de la siguiente sección.

Recuadro de aplicación

Time Inc., la editorial de revistas más grande de Estados Unidos, maneja un portafolio de más de 125 revistas. Uno de cada dos estadounidenses adultos lee una de las publicaciones de Time Inc. al mes.

Una revista es un buen ejemplo de un producto perecedero dada la rapidez con la que cada número pierde actualidad, por lo que el modelo de inventario que se describió en esta sección tiende a adaptarse muy bien a ellas. Desde la perspectiva de Time Inc., este "problema del voceador" de cada revista se presenta en tres niveles —corporativo, de mayorista y de menudeo— pero la complejidad de cada caso no refleja en los supuestos del modelo. A nivel corporativo se debe tomar una decisión acerca del número de ejemplares que se debe imprimir, pues la demanda de la revista se encuentra determinada en gran medida por las negociaciones con los mayoristas más que por una variable aleatoria. De forma similar, cada mayorista debe decidir cuántos ejemplares pedirá, lo que depende de las negociaciones con sus minoristas, más que de una variable aleatoria. Por su parte, la demanda de la revista por parte de cada minorista es, de hecho, una variable aleatoria, pero los datos necesarios para generar una estimación razonable de la distribución de probabilidad de la variable aleatoria pueden no estar disponibles. (Por ejemplo, si un determinado número de una revista se agota antes de que salga el siguiente, el minorista no podrá determinar cuál hubiera sido la demanda si dicho minorista hubiera contado con la cantidad suficiente de revistas.)

Con la ayuda de un consultor en IO se formó un grupo de trabajo sobre la *investigación de la administración de inventarios* con el fin de determinar cómo integrar mejor las decisiones que se toman en los tres niveles. El equipo comenzó a construir el modelo a partir del nivel inferior (el del minorista) y llevó a cabo un análisis de IO que permitiera hacer un mejor uso de los datos disponibles a fin de evaluar cada orden de impresión de la revista y los procedimientos de distribución tanto del mayorista como del minorista. Las soluciones bien conocidas de modelos formales de inventarios tuvieron que ser adaptados de tal forma que pudieran implantarse dentro de las restricciones del canal de distribución de la revista. A pesar de ello, dicho estudio tuvo éxito en desarrollar un nuevo y bien diseñado proceso de distribución de tres escalones. La adopción de este nuevo proceso ha resultado en *la generación creciente de ganancias del orden de más de* **3.5 millones de dólares** *anuales* para Time Inc.

Fuente: M. A. Koschat, G. L. Berk, J. A. Blatt, N. M. Kunz, M. H. LePore y S. Blyakher: "Newsvendors Tackle the Newsvendor Problem", en *Interfaces*, **33**(3): 72-84, mayo-junio de 2003. (En el sitio en internet de este libro —www.mhhe.com/hillier— se proporciona una liga hacia este artículo).

En ocasiones, cuando se administra el inventario de estos tipos de productos perecederos, es necesario estudiar algunas cuestiones, además de las que se presentarán en esta sección. Se ha realizado una investigación extensa para ampliar el modelo de manera que cubra estas consideraciones y se ha logrado un avance importante. (En la referencia seleccionada 5 se proporciona una revisión bibliográfica de esta investigación.)[10]

Ejemplo

En este caso se hace referencia al ejemplo 2 de la sección 18.1, que involucra la distribución de un modelo específico de bicicleta. Ahora existen otras condiciones. El fabricante acaba de informar al distribuidor que este modelo se va a descontinuar. Además, para ayudar a agotar su inventario, le ofrece al distribuidor la oportunidad de hacer una compra final en términos muy favorables, esto es, por un *costo unitario* de $200 cada bicicleta. Con este arreglo especial, el distribuidor *no* incurre en un *costo de preparación* al colocar esta orden.

El distribuidor cree que esta oferta le proporciona una oportunidad ideal para hacer una venta final a sus clientes (tiendas de bicicletas) para la temporada de Navidad a un precio reducido de 450 dólares por unidad, esto es, con una ganancia de 250 dólares por cada una. Ésta debe ser una venta de una sola vez porque el modelo se sustituirá por uno nuevo que lo hará caer en la obsolescencia. Entonces, cualquier bicicleta que no venda durante esta barata prácticamente no tendrá valor. Sin embargo, el distribuidor cree que podrá disponer de las bicicletas que queden después de Navidad si las vende a un precio nominal de 100 dólares cada una (*valor de rescate*) y así recupera la mitad del precio de compra. Considerando esta pérdida por ordenar más de lo que puede vender, lo mismo que de la ganancia perdida si ordena menos de lo que puede vender, el distribuidor debe decidir qué cantidad ordenará al fabricante.

El costo administrativo en el que se incurre al colocar órdenes especiales para la temporada navideña es bastante pequeño, por lo que este costo será pasado por alto hasta casi el final de esta sección.

[10] Dentro de las investigaciones más recientes se encuentra G. Raz y E. L. Porteus, "A Fractiles Perspective to the Joint Price/Quantity Newsvendor Model", en *Management Science*, **52**(11): 1764-1777, noviembre de 2006.

Otro gasto relevante es el costo de mantener las bicicletas no vendidas en inventario hasta que pueda disponerse de ellas después de Navidad. Al combinar el costo de capital comprometido en el inventario y otros costos de almacenamiento, este costo de inventario se estima en 10 dólares por bicicleta que queda después de Navidad. En consecuencia, si el valor de rescate es de 100 dólares, el *costo unitario de mantener* es –90 dólares por bicicleta que queda en inventario al final.

Es necesario estudiar otros dos componentes de costo: el costo por faltantes y el ingreso. Si la demanda excede los recursos, los clientes que no pudieron comprar una bicicleta quizá se molesten, lo que ocasionará un "costo" al distribuidor. Este costo por faltantes es la cuantificación, por cada artículo, de la pérdida de imagen multiplicada por la cantidad de demanda que no se satisface cuando hay faltantes. El distribuidor considera que este costo es despreciable.

Si se adopta el criterio de maximizar la ganancia, el modelo debe incluir el ingreso. La ganancia total es igual al ingreso total menos el costo en que se incurre (al ordenar, almacenar y tener faltantes). Bajo el supuesto que *no hay inventario inicial*, la ganancia para el distribuidor es

$$\begin{aligned}\text{Ganancia} = {} & \$450 \times \text{cantidad vendida por el distribuidor} \\ & - \$200 \times \text{cantidad comprada por el distribuidor} \\ & + \$90 \times \text{cantidad no vendida y liquidada a valor de rescate.}\end{aligned}$$

Sea

$$\begin{aligned}S &= \text{cantidad comprada al distribuidor} \\ &= \text{nivel de inventario después de recibir esta compra (puesto que no hay inventario inicial)}\end{aligned}$$

y

$$D = \text{demanda de las tiendas de bicicletas (variable aleatoria),}$$

de manera que

$$\begin{aligned}\text{mín}\{D, S\} &= \text{cantidad vendida,} \\ \text{máx}\{0, S - D\} &= \text{cantidad no vendida.}\end{aligned}$$

Entonces

$$\text{Ganancia} = 450 \text{ mín}\{D, S\} - 200s + 90 \text{ máx}\{0, S - D\}.$$

El primer término también se puede escribir como

$$450 \text{ mín}\{D, S\} = 450D - 450 \text{ máx}\{0, D - S\}.$$

El término $450 \text{ máx}\{0, D - S\}$ representa el *ingreso perdido por la demanda insatisfecha*. Este ingreso perdido más cualquier costo por la pérdida de imagen ante los clientes debida a la demanda insatisfecha (que se supuso despreciable en este ejemplo) se interpretará como el *costo por faltantes* en esta sección.

Ahora observe que $450D$ es independiente de la política de inventario (el valor de S seleccionado) y, por lo tanto, se puede eliminar de la función objetivo, lo que deja

$$\text{Ganancia relevante} = -450 \text{ máx}\{0, D - S\} - 200S + 90 \text{ máx } \{0, S - D\}$$

para maximizar. Todos los términos del lado derecho son los *negativos* de los *costos*, que representan el *costo por faltantes*, el *costo de ordenar* y el *costo de mantener* (que en este caso tiene un valor negativo), respectivamente. En lugar de *maximizar* el *negativo* del costo total, se *minimizará* la función equivalente

$$\text{Costo total} = 450 \text{ máx}\{0, D - S\} + 200S - 90 \text{ máx}\{0, S - D\}$$

De manera más precisa, como el costo total es una variable aleatoria (ya que D es una variable aleatoria), el objetivo adoptado para el modelo es *minimizar el costo total esperado*.

En el análisis sobre la interpretación del costo por faltantes se supuso que la demanda insatisfecha se perdía (no se surte después). En el caso que esta demanda insatisfecha se cumpla

mediante un envío prioritario, se aplica un razonamiento similar. El componente del rendimiento del ingreso neto se convierte en el precio de venta de una bicicleta (450 dólares) multiplicado por la demanda *menos* el costo unitario del envío prioritario multiplicado por la demanda insatisfecha cuando ocurre un faltante. Si el distribuidor se ve forzado a cumplir con la demanda insatisfecha mediante la compra de las bicicletas a otro distribuidor en 350 dólares cada una más el cargo por envío aéreo de, por ejemplo, 20 dólares por cada una, entonces el costo por faltantes apropiado es de 370 dólares por bicicleta. (En caso de haber algún costo asociado por la pérdida de imagen, también se debe agregar a esta cantidad.)

El distribuidor no sabe cuál será la demanda de bicicletas, es decir, la demanda D es una variable aleatoria. Sin embargo, se puede elaborar una política óptima de inventario si se dispone de información sobre la distribución de probabilidad de D. Sea

$$P_D(d) = P\{D = d\}.$$

Se supondrá que $P_D(d)$ se conoce para todos los valores de $d = 0, 1, 2, \ldots$

Ahora se cuenta con los elementos para resumir el modelo en términos generales.

Supuestos del modelo

1. Cada aplicación incluye un solo producto perecedero.
2. Cada aplicación incluye un solo periodo porque el producto no se puede vender después.
3. No obstante, será posible disponer de las unidades del producto que queden al final del periodo, quizá incluso con *valor de rescate* por las unidades.
4. Puede haber algún inventario inicial al comienzo de este periodo, denotado por

$$I = \text{inventario inicial.}$$

5. La única decisión que debe tomarse es el número de unidades que es necesario ordenar (para compra o bien para producción) de manera que se pueden colocar en el inventario al principio del periodo. Así

$$\begin{aligned} Q &= \text{cantidad por ordenar,} \\ S &= \text{nivel de inventario después de recibir esta orden} \\ &= I + Q. \end{aligned}$$

   Dado I, será conveniente usar S como la variable de decisión del modelo, la cual determina entonces de manera automática $Q = S - I$.
6. La *demanda* para retirar unidades del inventario y venderlas (o para algún otro propósito) durante el periodo es una variable aleatoria D. Sin embargo, se conoce la distribución de probabilidad de D (o al menos se puede estimar).[11]
7. Después de eliminar el ingreso si se satisface la demanda (puesto que ésta es independiente de la decisión S), el objetivo se convierte en minimizar el costo total esperado, cuyos componentes son

$K =$ costo de preparación para comprar o producir el lote completo de unidades,
$c =$ costo unitario de comprar o producir cada unidad,
$h =$ costo de mantener cada unidad que queda al final del periodo (incluye el costo de almacenar menos el valor de rescate),
$p =$ costo por faltantes por unidad de demanda no satisfecha (incluye el ingreso perdido y el costo de la pérdida de imagen ante el cliente).

[11] En la práctica, es necesario estimar la probabilidad de distribución a partir de una cantidad limitada de datos pasados acerca de la demanda. Entre las investigaciones que se han llevado a cabo acerca de cómo eliminar el supuesto 6 y de cómo aplicar los datos disponibles acerca de la demanda directamente se destacan R. Levi, R. O. Roundy y D. B. Shmoys, "Provably Near-Optimal Sampling-Based Policies for Stochastic Inventory Control Models", en *Mathematics of Operations Research*, **32**(4): 821-839, noviembre de 2007. También consulte L. Y. Chu, J.G. Shanthikumar y Z.-J. M. Shen, "Solving Operational Statistics Via a Bayesian Analysis", en *Operations Research Letters*, **36**(1): 110-116, enero de 2008.

Análisis del modelo sin inventario inicial ($I = 0$) y sin costo de preparación ($K = 0$)

Antes de analizar el modelo en toda su generalidad, será provechoso comenzar por considerar el caso más simple en el que $I = 0$ (no hay inventario inicial) y $K = 0$ (no hay costo de preparación).

La decisión sobre el valor de S, la cantidad de inventario que se debe adquirir, depende en gran medida de la distribución de probabilidad de la demanda D. Quizá sea deseable superar la demanda esperada, pero tal vez sin alcanzar la demanda máxima posible. Es necesario una compensación entre 1) el riesgo de una escasez que implica incurrir en costos por faltantes y 2) el riesgo de tener un excedente e incurrir en costos de desperdicio por ordenar y almacenar unidades en exceso. Este objetivo se logra mediante la minimización del valor esperado (en el sentido estadístico) de las sumas de estos costos.

La cantidad vendida está dada por

$$\text{mín}\{D, S\} = \begin{cases} D & \text{si } D < S \\ S & \text{si } D \geq S. \end{cases}$$

En consecuencia, si la demanda es D y se almacena S, el costo en que se incurre está dado por

$$C(D, S) = cS + p \text{ máx}\{0, D - S\} + h \text{ máx}\{0, S - D\}.$$

Como la demanda es una variable aleatoria [con distribución de probabilidad $P_D(d)$], este costo también es una variable aleatoria. El costo esperado está dado por $C(S)$, donde

$$\begin{aligned} C(S) = E[C(D, S)] &= \sum_{d=0}^{\infty} (cS + p \text{ máx}\{0, d - S\} + h \text{ máx}\{0, S - d\})P_D(d) \\ &= cS + \sum_{d=S}^{\infty} p(d - S)P_D(d) + \sum_{d=0}^{S-1} h(S - d)P_D(d). \end{aligned}$$

La función $C(S)$ depende de la distribución de probabilidad de D. Con frecuencia es difícil encontrar una representación de ella, en particular cuando la demanda tiene un gran número de valores posibles. Por ello, muchas veces esta *variable aleatoria discreta* se aproxima por una *variable aleatoria continua*. Más aún, cuando la demanda tiene un gran número de valores posibles, esta aproximación casi siempre conducirá a un valor muy cercano a la cantidad óptima de inventario. Además, cuando se usa una demanda discreta, la solución analítica de las expresiones puede resultar más difícil. Entonces, a menos que se establezca de otra manera, se supondrá una *demanda continua* en lo que queda del capítulo.

En el caso de esta variable aleatoria D, sea

$$f(x) = \text{función de densidad de probabilidad de } D$$

y

$$F(d) = \text{función de distribución acumulada (FDA) de } D,$$

de manera que

$$F(d) = \int_0^d f(x)\, dx.$$

Cuando se selecciona un nivel de inventario por ordenar S, la FDA $F(d)$ se convierte en la probabilidad de que *no* ocurrirá un faltante antes de que termine el periodo. Igual que en la sección anterior, esta probabilidad se conoce como el **nivel de servicio** que proporciona la cantidad que se ordenará. El costo esperado $C(S)$ correspondiente se expresa como

$$\begin{aligned} C(S) = E[C(D, S)] &= \int_0^{\infty} C(x, S)f(x)\, dx \\ &= \int_0^{\infty} (cS + p \text{ máx}\{0, x - S\} + h \text{ máx}\{0, S - x\})f(x)\, dx \\ &= cS + \int_S^{\infty} p(x - S)f(x)\, dx + \int_0^S h(S - x)f(x)\, dx. \end{aligned}$$

Ahora es necesario obtener el valor de S, sea S^*, que minimiza $C(S)$. Encontrar la fórmula de S^* requiere una derivación relativamente compleja, por lo que aquí sólo se dará el resultado. Sin embargo, la derivación se proporciona en en el sitio en internet de este libro como un complemento a este capítulo para el lector con una inclinación más matemática. (Este complemento también extiende brevemente el modelo al caso en el que los costos de mantener y costos de faltantes son funciones *no lineales* en lugar de lineales.)

En este complemento se muestra que la función $C(S)$ tiene una forma parecida a la que se presenta en la figura 18.14, porque es una función *convexa* (es decir, la segunda derivada es *no negativa* en cualquier punto). En realidad, se trata de una función *estrictamente convexa* (es decir, la segunda derivada es *estrictamente positiva*) en cualquier punto si $f(x) > 0$ para toda $x \geq 0$. Además, la primera derivada se vuelve positiva ante una S suficientemente grande, así que $C(S)$ debe poseer un mínimo global. Este mínimo global se muestra en la figura 18.14 como S^*, por lo cual $S = S^*$ es el nivel de inventario óptimo que se debe obtener cuando se recibe la cantidad ordenada ($Q = S^*$) al inicio del periodo.

En particular, el complemento determina que el nivel de inventario óptimo S^* es el valor que satisface

$$F(S^*) = \frac{p - c}{p + h}.$$

En consecuencia, $F(S^*)$ es el *nivel de servicio óptimo* y la cantidad correspondiente S^* se puede obtener ya sea mediante la solución algebraica de esta ecuación o la gráfica de la FDA donde se identifica S^*. Para interpretar el lado derecho de esta ecuación, el numerador se puede ver como

$p - c$ = costo unitario de ordenar menos
= disminución de la ganancia debida a que no se ordena una unidad que se pudo haber vendido durante ese periodo.

De manera similar,

$c + h$ = costo unitario de ordenar más
= disminución de la ganancia debida por ordenar una unidad que no se pudo vender durante ese periodo.

Por lo tanto, si se denota el costo unitario de ordenar menos o más por C_{menos} y $C_{\text{más}}$, respectivamente, esta ecuación especifica que

$$\text{Nivel de servicio óptimo} = \frac{C_{\text{menos}}}{C_{\text{menos}} + C_{\text{más}}}.$$

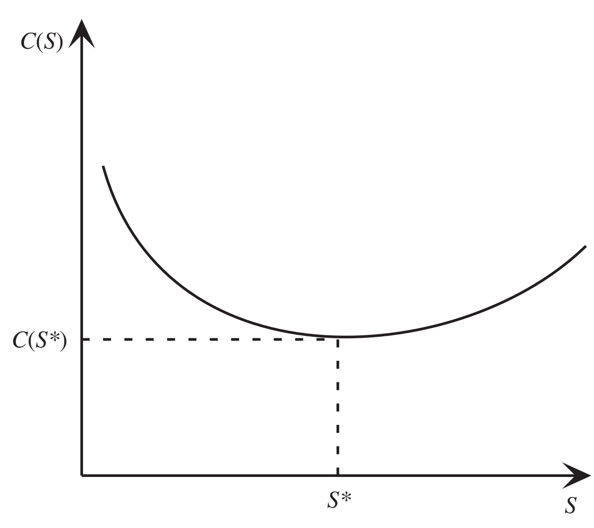


FIGURA 18.14
Gráfica de $C(S)$, costo esperado según el modelo estocástico de un solo periodo de productos perecederos como una función de S (el nivel de inventario cuando la cantidad por ordenar $Q = S - I$ se recibe al comienzo del periodo), dado que el inventario inicial es $I = 0$ y el costo de preparación es $K = 0$.

En el IOR Tutorial se proporciona una plantilla de Excel para calcular S^*, cuando la demanda tiene distribución uniforme o distribución exponencial. También se incluye una plantilla similar de Excel en los archivos de Excel de este capítulo dentro del sitio en internet de este libro.

Si se supone que D es una variable aleatoria discreta con distribución acumulada

$$F(d) = \sum_{n=0}^{d} P_D(n),$$

se obtiene un resultado similar de la cantidad óptima por ordenar. En particular, la cantidad óptima por ordenar, S^*, es el entero más pequeño tal que

$$F(S^*) \geq \frac{p - c}{p + h}.$$

En la sección Worked Examples del sitio en internet de este libro se proporciona un ejemplo que implica la sobreventa en aerolíneas donde D es una variable aleatoria discreta. El ejemplo que se presenta a continuación trata a D como una variable aleatoria continua.

Aplicación al ejemplo

De nuevo en el caso del ejemplo de las bicicletas que se describió al inicio de esta sección, suponga que la demanda tiene una distribución exponencial con media de 10 000, de manera que su función de densidad de probabilidad es

$$f(x) = \begin{cases} \dfrac{1}{10\,000} e^{-x/10\,000} & \text{si } x \geq 0 \\ 0 & \text{de otra forma} \end{cases}$$

y la función FDA es

$$F(d) = \int_0^d \frac{1}{10\,000} e^{-x/10\,000}\, dx = 1 - e^{-d/10\,000}.$$

A partir de los datos proporcionados,

$$c = 200, \qquad p = 450, \qquad h = -90.$$

En consecuencia, S^* (el nivel óptimo de inventario que se debe alcanzar para comenzar a satisfacer la demanda) es el valor que cumple con

$$1 - e^{-S^*/10\,000} = \frac{450 - 200}{450 - 90} = 0.69444.$$

Si se toma el logaritmo natural (denotado por ln), esta ecuación se resuelve como sigue:

$$\begin{aligned} e^{-S^*/10\,000} &= 0.30556, \\ \ln e^{-S^*/10\,000} &= \ln 0.30556, \\ \frac{-S^*}{10\,000} &= -1.1856, \\ S^* &= 11\,856. \end{aligned}$$

Por ello, el distribuidor debe almacenar 11 856 bicicletas en la temporada de Navidad. Note que este número es un poco mayor que la demanda esperada de 10 000.

Siempre que la demanda tenga una distribución exponencial con un valor esperado de λ, S^* se puede obtener a partir de la relación

$$S^* = -\lambda \ln \frac{c + h}{p + h}.$$

Análisis del modelo con inventario inicial ($I > 0$) pero sin costo de preparación ($K = 0$)

Ahora considere el caso en el que $I > 0$, es decir, que existen I unidades en inventario dentro del periodo pero antes de recibir una orden por cierta cantidad, $Q = S - I$. (Por ejemplo, este caso surgiría para el ejemplo de las bicicletas si el distribuidor comienza con 500 bicicletas antes de colocar una orden, por lo que $I = 500$.) Se continúa con el supuesto de que $K = 0$ (no hay costo de preparación).

Sea

$\overline{C}(S)$ = costo esperado del modelo para cualquier valor de I y K (incluye el supuesto actual de que $K = 0$), dado que S es el nivel de inventario que se obtiene cuando se recibe la cantidad ordenada al inicio del periodo,

Entonces el objetivo es escoger una $S \geq I$ para

$$\underset{S \geq I}{\text{Minimizar}} \quad \overline{C}(S).$$

Resulta instructivo comparar $\overline{C}(S)$ con la función de costo que se utilizó en la sección anterior (y que se grafica en la figura 18.14),

$C(S)$ = costo esperado del modelo, dado S, cuando $I = 0$ y $K = 0$.

Con $K = 0$,

$$\overline{C}(S) = c(S - I) + \int_S^{\infty} p(x - S)f(x)dx + \int_0^S h(S - x)f(x)dx.$$

En consecuencia, $\overline{C}(S)$ es idéntico a $C(S)$ excepto por el primer término, donde $C(S)$ tiene cS en vez de $c(S - I)$. Por lo tanto,

$$\overline{C}(S) = C(S) - cI.$$

Como I es una constante, esto significa que $\overline{C}(S)$ alcanza su mínimo en el mismo valor de S^* que para $C(S)$, como se muestra en la figura 18.14. Sin embargo, como S debe estar restringida a $S \geq I$, si $I > S^*$ la figura 18.14 indica que $\overline{C}(S)$ se minimizaría cuando $S \geq I$ al establecer $S = I$ (es decir, al no colocar una orden). Esta situación produce la siguiente política de inventarios.

Política óptima de inventarios con $I > 0$ y $K = 0$

Si $I < S^*$, se ordena $S^* - I$ para subir el nivel de inventario a S^*.
Si $I \geq S^*$, no se ordena,

donde S^* de nuevo satisface

$$F(S^*) = \frac{p - c}{p + h}.$$

Por consiguiente, en el ejemplo de las bicicletas, si se tienen 500 unidades en inventario, la política óptima será ordenar hasta 11 856 (lo que significa ordenar 11 356 bicicletas adicionales). Por otro lado, si ya se tienen 12 000 bicicletas en inventario, la política óptima será no ordenar.

Análisis del modelo con un costo de preparación ($K > 0$)

Ahora considere la versión restante del modelo donde $K > 0$, en cuyo caso se incurre en un costo de preparación K por comprar o producir el lote completo de unidades que se ha ordenado. (En el ejemplo de las bicicletas, si se incurriera en un costo administrativo de 8 000 dólares por colocar la orden especial de bicicletas para la temporada navideña, entonces $K = 8\,000$.) Ahora se permitirá cualquier valor del inventario inicial, por lo tanto $I \geq 0$.

Con $K > 0$, el costo esperado $\overline{C}(S)$, dado el valor de la variable de decisión S, es

$$\overline{C}(S) = K + c(S - I) + \int_S^{\infty} p(x - S)\, f(x)dx + \int_0^S h(S - x)\, f(x)\, dx \qquad \text{si se ordena;}$$
$$\overline{C}(S) = \int_S^{\infty} p(x - S)\, f(x)\, dx + \int_0^S h\,(S - x)\, f(x)\, dx \qquad \text{si no se ordena.}$$

Por lo tanto, en comparación con la función de costo esperado $C(S)$ que se grafica en la figura 18.14 (donde se supone que $I = 0$ y $K = 0$),

$$\overline{C}(S) = K + C(S) - cI \qquad \text{si se ordena;}$$
$$\overline{C}(I) = C(I) - cI \qquad \text{si no se ordena.}$$

Como I es una constante, el término cI en ambas expresiones se puede pasar por alto para propósitos de minimizar $\overline{C}(S)$ cuando $S \geq I$. En consecuencia, la gráfica de $C(S)$ de la figura 18.14 se puede usar para determinar si se debe colocar una orden, y si es así, qué valor de S debería seleccionarse.

Esto es lo que se hace en la figura 18.15, donde s^* es el valor de S tal que

$$C(s^*) = K + C(S^*).$$

Por lo tanto,

si $I < s^*$, entonces $C(S^*) < K + C(I)$, por lo que se debe ordenar con $S = S^*$;
si $I \geq s^*$, entonces $C(S) \leq K + C(I)$ para cualquier $S \geq I$, así que no se debe ordenar

En otras palabras, si el inventario inicial I es menor que s^*, entonces el gasto que implica el costo de preparación K no tiene sentido porque al subir el nivel de inventario hasta S^* (cuando se ordena $S – I$) se reducirá el costo restante esperado en más de K cuando se compara con la opción de no ordenar. Sin embargo. Si $I > s^*$, se hace imposible recobrar el costo de preparación K independientemente de la cantidad que se ordene. (Si $I = s^*$, al incurrir en el costo de preparación K para ordenar $S^* – s^*$ se reducirá el costo restante esperado en la misma cantidad, por lo cual no existe razón para ordenar.) Lo anterior conduce a la siguiente política de inventarios.

Política de inventarios óptima con $I \geq 0$ y $K > 0$

Si $I < s^*$, ordenar $S^* – I$ para subir el nivel del inventario a S^*.
Si $I^* \geq s^*$, no ordenar.
(Vea la fórmula enmarcada de S^* y s^* que se dio con anterioridad.)

Cuando la demanda tiene una distribución uniforme o exponencial, s^* y S^* se pueden calcular por medio de una rutina automática que se presenta en el IOR Tutorial.

Este tipo de política es conocida como **política (*s*, *S*)**, la cual tiene un uso extendido en la industria.

La política (s, S) también se emplea con frecuencia cuando se aplican modelos estocásticos de revisión periódica a *productos estables*, de manera que se deben considerar periodos múltiples.

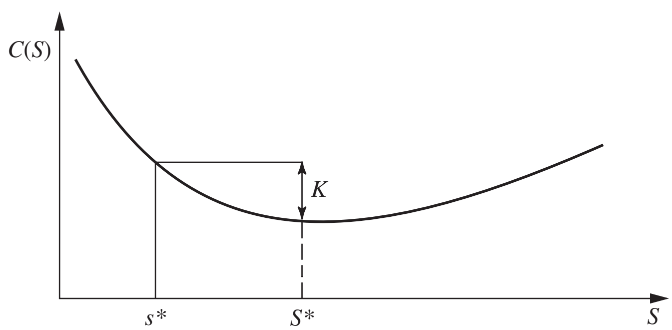


FIGURA 18.15
Gráfica de $C(S)$, costo esperado (dado S) según el modelo estocástico de un solo periodo cuando $I = 0$ y $K = 0$, que se usa aquí para determinar los puntos críticos s^* y S^* de la política de inventarios óptima en la versión del modelo donde $I \geq 0$ y $K > 0$.

En este caso resulta algo más complicado encontrar la política de inventarios óptima dado que es posible que los valores de s y S sean diferentes en periodos distintos. En el segundo complemento de este capítulo que se encuentra en el sitio en internet de este libro se proporcionan los detalles.

De regreso al modelo de un solo periodo, ahora se ilustrará el cálculo de la política de inventarios óptima para el ejemplo de las bicicletas cuando $K > 0$.

Aplicación al ejemplo

Suponga que el costo administrativo de colocar la orden especial de bicicletas durante la próxima temporada navideña se estima en 8 000 dólares. En consecuencia, ahora los parámetros del modelo son

$$K = 8\,000, \qquad c = 200, \qquad p = 450, \qquad h = -90.$$

Como ya se indicó, se supone que la demanda de bicicletas tiene una distribución exponencial con una media de 10 000 unidades.

Antes se encontró para este ejemplo que

$$S^* = 11\,856.$$

para determinar s^*, es necesario resolver la ecuación

$$C(s^*) = K + C(S^*),$$

para s^*. Si se sustituye dos veces en la expresión de $C(S)$ que se dio en la primera parte de esta sección, con $S = s^*$ en el lado izquierdo de la ecuación y $S = S^* = 11\,856$ en el lado derecho, la ecuación se convierte en

$$200s^* + 450\int_{s^*}^{\infty}(x - s^*)\frac{1}{10\,000}e^{-x/10\,000}dx - 90\int_{0}^{s^*}(s^* - x)\frac{1}{10\,000}e^{-x/10\,000}dx$$

$$= 8\,000 + 200(11\,856) + 450\int_{11\,856}^{\infty}(x - 11\,856)\frac{1}{10\,000}e^{-x/10\,000}dx$$

$$- 90\int_{0}^{11\,856}(11\,856 - x)\frac{1}{10\,000}e^{-x/10\,000}dx.$$

Después de realizar muchos cálculos para obtener el número del lado derecho y reducir el lado izquierdo a una expresión más simple en términos de s^*, finalmente esta ecuación conduce a la solución numérica

$$s^* = 10\,674.$$

Entonces, la política óptima indica que se debe elevar el nivel del inventario hasta $S^* = 11\,856$ bicicletas si la cantidad disponible es menor que $s^* = 10\,674$. De otra manera no se ordena.

Solución aproximada de política óptima cuando la demanda presenta una distribución exponencial

Como se acaba de ilustrar en este ejemplo, se requieren cálculos muy largos para encontrar s^* aun cuando la demanda tenga una distribución relativamente directa como la distribución exponencial. Por lo tanto, dada esta distribución, en seguida se desarrollará una aproximación cercana a la política de inventarios óptima que es fácil de calcular.

En la sección 17.4 se describió que, en el caso de una distribución exponencial con una media de $1/\alpha$, la función de densidad de probabilidad $f(x)$ y la FDA $F(x)$ son

$$f(x) = \alpha e^{-\alpha x}, \qquad \text{para } x \geq 0,$$
$$F(x) = 1 - e^{-\alpha x}, \qquad \text{para } x \geq 0.$$

En consecuencia, dado que

$$F(S^*) = \frac{p - c}{p + h},$$

se tiene

$$1 - e^{-\alpha S^*} = \frac{p - c}{p + h}, \qquad \text{o} \qquad e^{-\alpha S^*} = \frac{(p + h) - (p - c)}{p + h} = \frac{h + c}{h + p},$$

de manera que

$$S^* = \frac{1}{\alpha} \ln \frac{h + p}{h + c}$$

es la solución exacta de S^*.

Para comenzar a desarrollar una aproximación de s^*, se comienza con la ecuación exacta,

$$C(s^*) = K + C(S^*).$$

Puesto que

$$C(S) = cS + h \int_0^S (S - x)\alpha e^{-\alpha x}\, dx + p \int_S^\infty (x - S)\alpha e^{-\alpha x}\, dx$$
$$= (c + h)S + \frac{1}{\alpha} (h + p)e^{-\alpha S} - \frac{h}{\alpha}.$$

Esta ecuación se convierte en

$$(c + h)s^* + \frac{1}{\alpha}(h + p)e^{-\alpha s^*} - \frac{h}{\alpha} = K + (c + h)S^* + \frac{1}{\alpha}(h + p)e^{-\alpha S^*} - \frac{h}{\alpha},$$

o (si se usan los resultados anteriores de S^*)

$$(c + h)s^* + \frac{1}{\alpha}(h + p)e^{-\alpha s^*} = K + (c + h)S^* + \frac{1}{\alpha}(c + h).$$

Aunque esta última ecuación no tiene una forma de solución corta de s^*, su solución numérica es bastante sencilla. Una solución analítica aproximada se puede obtener de la manera siguiente. Sea

$$\Delta = S^* - s^*,$$

si se observa que

$$e^{-\alpha S^*} = \frac{h + c}{h + p},$$

esta ecuación conduce a

$$\frac{1}{\alpha}(h + p) \frac{e^{-\alpha s^*}}{e^{-\alpha S^*}} = \frac{K + (c + h)\Delta + \frac{1}{\alpha}(c + h)}{\frac{h + c}{h + p}},$$

lo cual se reduce a

$$e^{\alpha\Delta} = \frac{\alpha K}{c + h} + \alpha\Delta + 1.$$

Si $\alpha\Delta$ es cercano a cero, $e^{\alpha\Delta}$ se puede expandir en una serie de Taylor alrededor de cero. Si se desprecian los términos más allá del término cuadrático, el resultado se convierte en

$$1 + \alpha\Delta + \frac{\alpha^2\Delta^2}{2} \cong \frac{\alpha K}{c + h} + \alpha\Delta + 1,$$

de manera que

$$\Delta \cong \sqrt{\frac{2K}{\alpha(c+h)}}.$$

Por lo tanto, la aproximación deseada de s^* es

$$s^* \cong S^* - \sqrt{\frac{2K}{\alpha(c+h)}}.$$

Si se usa esta aproximación en el ejemplo de las bicicletas se obtiene

$$\Delta \cong \sqrt{\frac{(2)(10\,000)(800)}{200-90}} = 1\,206,$$

de manera que

$$s^* \cong 11\,856 - 1\,206 = 10\,650,$$

que es bastante cercano al valor exacto de $s^* = 10\,674$.

18.8 ADMINISTRACIÓN DE LOS INGRESOS

Al comienzo de la sección anterior se incluyó una lista de 12 ejemplos de productos perecederos. El último de ellos (el de las reservaciones proporcionadas por una línea aérea para hacer el inventario de asientos disponibles en un vuelo en particular) es de gran interés histórico debido a que su análisis condujo a un área de aplicación muy amplia y exitosa de la investigación de operaciones llamada comúnmente *administración de los ingresos*.

El punto de arranque de la administración de los ingresos fue la Airline Deregulation Act de 1978, el cual liberaba el control de los precios de las tarifas de las líneas aéreas. En esa época, varias líneas aéreas nuevas y empresas de flete ingresaron al mercado. Una de las más importantes, American Airlines, encabezó la guerra de precios cuando introdujo las *tarifas de descuento con capacidad controlada*. En varios vuelos se vendieron un número limitado de asientos con descuento a medida que se necesitaba el servicio con el fin de igualar o mejorar las tarifas ofrecidas por las líneas aéreas de bajo costo, pero con restricciones que imponían que la compra debía realizarse con mucha anticipación (en un inicio, 30 días) al día de la salida. Las tarifas normales se seguían ofreciendo a los viajeros de negocios, quienes por lo general hacen sus reservaciones aun después de la fecha en la que vencen las tarifas con descuento. (El primer modelo en esta sección trata acerca de esta situación.)

Otra de las prácticas más antiguas y exitosas de la administración de los ingresos en la industria aeronáutica ha sido la del *sobrecupo* (ofrecer más reservaciones que el número de asientos disponibles con el fin de contrarrestar la gran cantidad de pasajeros que pierden el vuelo, lo que es algo normal). La regla de oro en la industria es que aproximadamente 15% de todos los asientos en un vuelo pueden ir desocupados si no existe alguna forma de sobrecupo. Por lo tanto, puede obtenerse un gran ingreso adicional si se recurre a la práctica del sobrecupo pero sin incurrir en el riesgo de sobrevender el vuelo. Sin embargo, las sanciones por negar el acceso a un vuelo a una persona con reservación son cuantiosas, por lo que se debe hacer un análisis cuidadoso para lograr una compensación adecuada entre la ganancia adicional que genera sobrevender un vuelo y el riesgo que implica cometer estas faltas. (El segundo modelo de esta sección trata de esta situación.)

Cuando se implanta la administración de los ingresos, las líneas aéreas de gran tamaño deben procesar reservaciones para muchas decenas de miles de pasajeros que viajan diariamente. Por lo tanto, mientras los modelos y algoritmos de IO se utilizan para administrar los ingresos, el otro componente esencial es una compleja tecnología de información. Por fortuna, los avances en el área de la tecnología de información de los años 80 proporcionaron las facilidades necesarias para automatizar las transacciones, capturar y almacenar grandes cantidades de datos, ejecutar a gran velocidad algoritmos complejos e implantar y administrar, con gran detalle, decisiones de administración de los ingresos.

En 1990, la práctica de la administración de los ingresos de American Airlines se había refinado a tal punto que generaba alrededor de 500 millones de dólares de ingresos adicionales anualmente. (La referencia seleccionada A8 narra la historia.) Para entonces, otras aerolíneas también empezaron a ver la forma de desarrollar sistemas similares de administración de los ingresos.

Como resultado de esta historia, la práctica de la administración de los ingresos en la industria aeronáutica ha tenido una gran penetración, pues ha sido altamente desarrollada y muestra una gran eficiencia. De acuerdo con la página 10 de la referencia seleccionada 10 (el tratado oficial acerca de la teoría y práctica de la administración de los ingresos), "la mayoría de las estimaciones, las ganancias producto del uso de los sistemas de administración de los ingresos son *groso modo* comparables a la rentabilidad total de muchas aerolíneas en un buen año (alrededor de 4 a 5% de ganancias)".

El enorme éxito que obtuvo la industria aeronáutica en esta área ha llevado a que otros sectores de servicios con características similares desarrollen sus propios sistemas de administración de los ingresos. De entre ellos se destacan el hotelero, las líneas de cruceros, los ferrocarriles de pasajeros, las compañías de renta de autos, los operadores de viajes, los teatros y los estadios deportivos. La administración de los ingresos está en crecimiento en la industria minorista en el manejo de productos rápidamente perecederos (por ejemplo, en las tiendas misceláneas), en productos estacionales (por ejemplo, tiendas de ropa) y productos que caen en la obsolescencia con rapidez (por ejemplo, tiendas de alta tecnología).

Para alcanzar dichos resultados sorprendentes a veces se requiere desarrollar sistemas de administración de los ingresos relativamente complejos con muchas categorías de clientes, tarifas que cambian con el tiempo, etc. Los modelos y algoritmos necesarios para soportar dichos sistemas son relativamente complejos y, por ende, están fuera del alcance de este libro. Sin embargo, para exponer la idea general, a continuación presentamos dos modelos básicos de tipos elementales de administración de los ingresos. Los componentes de cada modelo se describen en términos generales para que puedan adaptarse a cualquier tipo de compañía, pero, con la finalidad de ser más precisos, se menciona el contexto de las líneas aéreas. Cada modelo es seguido de un ejemplo de aerolínea.

Modelo de tarifas de descuento controladas por capacidad

Una compañía tiene un inventario con un cierto producto perecedero (como, por ejemplo, los asientos en un vuelo de una aerolínea) para vender a dos clases de clientes (tales como los viajeros por placer y los de negocios). Los clientes clase 2 llegan primero a comprar unidades independientes del producto a un precio con descuento que está diseñado con el fin de garantizar que todo el inventario puede venderse antes de que el producto caduque. Existe una fecha límite para solicitar el precio de descuento, pero la compañía puede declarar terminada la venta especial en cualquier momento, cuando crea que se ha vendido lo suficiente. Después de que la oferta del precio con descuento caduca, los clientes clase 1 comienzan a comprar unidades del producto a su costo normal. Se supone que se conoce la distribución de probabilidad de la demanda de los clientes clase 1. La decisión que debe tomarse es qué cantidad del inventario total debe reservarse para los clientes clase 1, de tal forma que el precio con descuento pueda ser descontinuado antes si el inventario que resta desciende a este nivel antes de que llegue la fecha límite anunciada para este descuento.

Los parámetros (y la variable aleatoria) del modelo son

L = tamaño del inventario de los productos perecederos disponibles para venta,
p_1 = precio por unidad que pagan los clientes clase 1,
p_2 = precio por unidad que pagan los clientes clase 2, donde $p_2 < p_1$,
D = demanda de los clientes clase 1 (variable aleatoria),
$F(x)$ = función de distribución acumulativa de D, por lo que $F(x) = P(D \leq x)$.

La variable de decisión es

x = nivel de inventario que debe reservarse para los clientes clase 1.

La clave para resolver el valor óptimo de x, expresada mediante x^*, radica en hacer la siguiente pregunta y después resolverla mediante el *análisis marginal*.

Pregunta: Suponga que en el inventario permanecen x unidades antes de la fecha de caducidad para solicitar el precio con descuento p_2 y un cliente clase 2 llega al mostrador deseando comprar una unidad a ese precio. ¿Tendrá que aceptarse o rechazarse esta solicitud?

Para resolver esta pregunta, es necesario comparar el ingreso incremental (o la esperanza estadística de los ingresos incrementales) de las dos opciones.

Si se acepta la solicitud, el ingreso incremental $= p^2$.

Si se rechaza la solicitud, el ingreso incremental $= \begin{cases} 0, & \text{si } D \leq x - 1 \\ p_1, & \text{si } D \geq x \end{cases}$

por lo que

$$E\text{ (ingreso incremental)} = p_1\, P(D \geq x)$$

Por lo tanto, la solicitud de hacer la venta al cliente clase 2 debe aceptarse si

$$p_2 > p_1\, P(D \geq x)$$

y rechazarse en cualquier otro caso. Ahora observe que $P(D \geq x)$ disminuye a medida que x aumenta. Por lo tanto, si la desigualdad es válida para un valor particular de x, dicho valor puede incrementarse hasta llegar al punto crítico x^* donde

$$p_2 \leq p_1\, P(D \geq x^*) \qquad \text{y} \qquad p_2 > p_1\, P(D \geq x^* + 1).$$

De aquí se deduce que el nivel de inventario óptimo para reservar a los clientes clase 1 es x^*. O, expresado de forma equivalente, el número máximo de unidades que deben venderse a los clientes clase 2 antes de descontinuar el precio con descuento p_2 es $L - x^*$.

Hasta el momento, se ha supuesto que los clientes compran unidades independientes del producto (tales como asientos en el vuelo de una aerolínea), por lo que la distribución de probabilidad de D podría ser una distribución discreta. Sin embargo, cuando el valor de L es grande (como en el caso del número de asientos en un vuelo lleno), puede ser mucho más conveniente desde el punto de vista computacional utilizar una distribución continua como una aproximación. También existen productos perecederos de los cuales se pueden adquirir cantidades fraccionarias, de tal forma que las distribuciones de demanda continua podrían ser las más adecuadas en este caso. Si se supone una distribución de demanda continua, al menos como aproximación, se puede deducir del análisis anterior que el nivel de inventario óptimo x^* que debe reservarse para los clientes clase 1 es el que satisface la ecuación,

$$p_2 = p_1\, P(D > x^*).$$

Puesto que $P(D > x^*) = 1 - P(D \leq x^*) = 1 - F(x^*)$, esta ecuación también puede escribirse como

$$F(x^*) = 1 - \frac{p_2}{p_1}.$$

(Cuando se utiliza una distribución continua como aproximación pero x^* que resuelve esas dos ecuaciones no es un entero, x^* deberá redondearse al entero inmediato *inferior* con la finalidad de satisfacer las expresiones que definen el valor entero óptimo de x^* que se dio al final del párrafo anterior.) Esta última ecuación muestra en forma clara que el cociente de p_2 entre p_1 juega un papel crítico en la determinación de la probabilidad de que se satisfará la totalidad de la demanda de los clientes clase 1.

Ejemplo que aplica este modelo a las tarifas de descuento controladas por capacidad

BLUE SKIES AIRLINES ha decidido aplicar este modelo en uno de sus vuelos. Este vuelo puede aceptar 200 reservaciones de asientos en la cabina principal. (Este número incluye una tolerancia para sobreventa ya que siempre hay pasajeros que no se presentan en el momento de la partida.)

El vuelo representa algo atractivo para un gran número de personas de negocios, quienes por lo general realizan sus reservaciones la víspera del viaje, pero que están dispuestos a pagar una tarifa relativamente elevada, de 1 000 dólares, por tener esta flexibilidad. Sin embargo, es necesario que la gran mayoría de los pasajeros viajen por placer a fin de llenar el avión. Por lo tanto, para atraer un número suficiente de este tipo de viajeros, se les ofrece una tarifa de descuento muy baja de 200 dólares siempre y cuando hagan sus reservaciones con al menos 14 días de anticipación y cumplan con otras restricciones (por ejemplo, no puede haber devoluciones).

Según la terminología de este modelo, los clientes clase 1 son pasajeros de negocios y los clase 2 son pasajeros que realizan viajes de placer, por lo que los parámetros del modelo son

$$L = 200, \qquad p_1 = \$1\,000, \qquad p_2 = \$200.$$

Mediante el uso de datos acerca del número de reservaciones que solicitaron los clientes clase 1 en cada vuelo en el pasado, se estima que la distribución de probabilidad del número de reservaciones que dicha clase de cliente solicitó para cada vuelo en el futuro se aproxima a una distribución normal con una media $\mu = 60$ y una desviación estándar $\sigma = 20$. Por lo tanto, ésta es la distribución de la variable aleatoria D del modelo, donde $F(x)$ representa la distribución acumulativa de D. Para encontrar el valor de x^*, el número óptimo de espacios para reservar a los clientes clase 1, se utiliza la ecuación que proporciona el modelo,

$$F(x^*) = 1 - \frac{p_2}{p_1} = 1 - \frac{\$200}{\$1\,000} = 0.8.$$

Mediante el uso de la tabla correspondiente a la distribución normal del apéndice 5 se obtiene

$$x^* = \mu + K_{0.2}\,\sigma = 60 + 0.842(20) = 76.84.$$

Puesto que, en realidad, x^* debe ser un entero, se redondea al entero inmediato *inferior* (como lo especifica el modelo), esto es, al entero 76. Mediante la reservación de 76 lugares para los clientes que están dispuestos a pagar la tarifa de 1 000 dólares por una reservación unos días antes del vuelo, esto implica que $L - x^* = 124$ es el número máximo de reservaciones que deben venderse a la tarifa de descuento de 200 dólares antes de descontinuar esta tarifa, aun si este hecho se presenta antes de la fecha límite de 14 días antes del vuelo.

Modelo de sobreventa

Lo mismo que en el modelo anterior, en este caso se trata de una compañía que tiene un inventario de un cierto producto perecedero (por ejemplo, los asientos en un vuelo comercial) para vender a sus clientes. No se hace ninguna distinción entre las diferentes clases de clientes. Las unidades del inventario están disponibles sólo en un cierto momento, por lo que cada cliente compra una unidad luego de hacer una reservación para adquirirla en el tiempo designado. No existe posibilidad de devolución. Sin embargo, no todos los clientes que hacen una reservación llegan a tiempo para obtener sus unidades. A los clientes que no llegan a la hora designada se les conoce como *ausentes*.

Debido a que la compañía prevé que habrá una gran cantidad de ausentes, puede aumentar sus ganancias mediante la **sobreventa** (vender una cantidad mayor de reservaciones que el inventario disponible). Sin embargo, se debe tener especial cuidado de no excederse en la sobreventa ya que existe una probabilidad elevada de que se pueda incurrir en una *falta de inventario* (más demanda que oferta). La razón es que existe un costo asociado a la *falta de oferta* cada vez que un cliente con reservación llega a tiempo para adquirir una unidad del inventario después de que éste se ha saturado. Por ejemplo, en la industria aeronáutica, se incurre en un costo por *negar el derecho a abordar* el avión cada vez que un cliente con una reservación es *rechazado* (se le niega el acceso al vuelo), lo que implica un costo puede incluir la devolución del dinero del boleto, una compensación por las molestias y el costo de la pérdida de imagen de la compañía (la pérdida de vuelos en el futuro). En algunos casos, el costo por negar el acceso al avión a un cliente puede consistir en la compensación que se le otorga a un cliente que tiene un asiento pero está dispuesto a dárselo a otro al que se le ha negado el derecho a un asiento.

La pregunta fundamental que plantea este modelo de sobreventa es cuánta sobreventa debe haber de tal forma que se maximicen las ganancias esperadas de la compañía. El modelo se basa en los siguientes supuestos:

1. Los clientes, de manera independiente, hacen sus reservaciones por una unidad del inventario y, por lo tanto, tienen la misma probabilidad de llegar a tiempo para adquirirla.
2. Existe un ingreso neto constante por cada reservación aceptada.
3. Existe un costo fijo por falta de inventario cada vez que un cliente con una reservación llega a tiempo para adquirir una unidad del inventario y se encuentra con que éste se ha agotado.

Con base en estos supuestos, el modelo tiene los parámetros siguientes:

$p =$ probabilidad de que un cliente que haga la reservación de una unidad del inventario llega a tiempo para adquirirla.
$r =$ ganancia neta que se obtiene por cada reservación aceptada.
$s =$ costo de escasez por unidad de la demanda insatisfecha.
$L =$ tamaño del inventario disponible.

La variable de decisión de este modelo es

$n =$ número de clientes a los que se les puede hacer la reservación de una unidad de inventario,

por lo que

$n - L =$ cantidad permitida de sobreventa.

Dado el valor de n, la incógnita es cuántos de los n clientes con reservación de una unidad del inventario llegarán a tiempo para adquirir la unidad. En otras palabras, ¿cuál es la *demanda* de unidades retiradas del inventario? Esta variable aleatoria se expresa mediante

$D(n) =$ demanda de unidades retiradas del inventario.

Se puede deducir a partir de los supuestos que $D(n)$ tiene una *distribución binomial* con parámetro p, por lo que

$$P\{D(n) = d\} = \binom{n}{d} p^d (1 - p)^{n-d} = \frac{n!}{d!(n - d)!} p^d (1 - p)^{n-d},$$

donde $D(n)$ tiene una media np y una varianza $np(1 - p)$.

Una variable aleatoria muy relacionada que será de gran importancia en nuestro análisis es la *demanda insatisfecha* que se presentará cuando se les otorgue una reservación a n clientes. Esta variable aleatoria se expresa mediante $U(n)$, por lo que

$$U(n) = \text{demanda insatisfecha} = \begin{cases} 0, & \text{si } D(n) \leq L \\ D(n) - L, & \text{si } D(n) > L \end{cases}$$

y

$$E(U(n)) = \sum_{d=L+1}^{n} (d - L)\, P\{D(n) = d\}.$$

Se utilizará el *análisis marginal* (el análisis del efecto de aumentar el valor de la variable de decisión n en 1) para determinar el valor óptimo de n que maximice la ganancia esperada, por lo que necesitaremos conocer el efecto en $E(U(n))$ de incrementar el valor de n en 1. Comenzando con n reservaciones, el efecto de sumar una reservación más es la de sumar 1 a la demanda insatisfecha sólo si ambos eventos ocurren. Un evento necesario es que las n reservaciones originales resulten en la saturación de todo el inventario, es decir, $D(n) \geq L$, y el otro evento que se requiere es que al cliente que ha hecho la reservación adicional llegue a tiempo para adquirir una unidad del inventario. De otra forma, no existirá ningún efecto en la demanda insatisfecha. Como consecuencia,

$$\Delta E(U(n)) = E(U(n + 1)) - E(U(n)) = p\, P\{D(n) \geq L\}$$

El valor de $\Delta E(U(n))$ depende del valor de n puesto que $P\{D(n) \geq L\}$, la probabilidad de saturar el inventario, también depende de n, el número de reservaciones. Para $n < L$, $\Delta E(U(n)) = 0$, mientras

que $\Delta E(U(n))$ aumenta a medida que n también lo hace puesto que la probabilidad de saturar el inventario se incrementa a medida que el número de reservaciones crece.

La variable aleatoria final de interés es la *ganancia* de la compañía que se presentará cuando n clientes tengan su reservación. Esta variable aleatoria se representa mediante $P(n)$, por lo que

$$P(n) = \text{ganancia} = r\,n - s\,U(n)$$
$$E(P(n)) = r\,n - s\,E(U(n)),$$
$$\Delta E(P(n)) = E(P(n+1)) - E(P(n)) = r - s\,\Delta E(U(n)) = r - s\,p\,P\{D(n) \geq L\}.$$

Como ya se mencionó, $\Delta E(U(n)) = 0$ para $n < L$, mientras que $\Delta E(U(n))$ aumenta a medida que n crece. Por lo tanto, $\Delta E(P(n)) > 0$ para valores relativamente pequeños de n y después (asumiendo que $r < s\,p$) cambiará a $\Delta E(P(n)) < 0$ para valores suficientemente grandes de n. Se deduce que n^*, el valor de n que maximiza $E(P(n))$, es el que satisface

$$\Delta E(P(n^* - 1)) > 0 \qquad \text{y} \qquad \Delta E(P(n^*)) \leq 0,$$

o, de forma equivalente,

$$r > s\,p\,P\{D(n^* - 1) \geq L\} \qquad \text{y} \qquad r \leq s\,p\,P\{D(n^*) \geq L\}.$$

Puesto que $D(n)$ tiene una distribución binomial, es un proceso directo (aunque tedioso desde el punto de vista computacional) obtener el valor de n^* de esta forma.

Cuando el valor de L es grande, es muy tedioso emplear la distribución binomial para llevar a cabo estos cálculos. Por lo tanto, es una práctica muy común usar la *aproximación normal* de la distribución binomial en esta aplicación (así como en muchas otras). En particular, la distribución normal con una media np y varianza $np(1 - p)$ con frecuencia se usa como una aproximación continua de la distribución binomial con parámetros n y p, puesto que esta última distribución tiene esta misma media y varianza. Mediante este método, podemos suponer que $D(n)$ tiene esta distribución normal y tratar a n como una variable de decisión continua. El valor óptimo de n está dado de manera aproximada mediante la ecuación

$$r = s\,p\,P\{D(n^* \geq L\}, \qquad \text{es decir,} \qquad P\{D(n^*) \geq L\} = \frac{r}{sp}$$

Usando la tabla de la distribución normal que se proporciona en el apéndice 5, el cálculo de n^* es directo, como lo ilustrará el ejemplo siguiente. Si n^* no es un entero, debe redondearse al entero inmediato superior a fin de satisfacer las expresiones que definen el valor entero óptimo de n^* que se proporciona al final del párrafo anterior.

Ejemplo de aplicación de este modelo de sobreventa

TRANSCONTINENTAL AIRLINES tiene un vuelo diario (excepto los fines de semana) de San Francisco a Chicago en el que principalmente vuelan personas de negocios. Hay 150 asientos disponibles en una sola cabina. La tarifa promedio por asiento es de 300 dólares. Esta tarifa no admite devoluciones, por lo que las personas ausentes pierden el monto total del boleto.

La política de la compañía es aceptar 10% más de reservaciones que el número de asientos disponibles en casi todos sus vuelos, ya que alrededor de ese porcentaje de todos sus clientes que hacen reservaciones no se presentan al vuelo. Sin embargo, si su experiencia con un vuelo en particular es muy diferente a este escenario, se puede hacer una excepción y se llama al grupo de IO para que analice cuál debe ser la política de sobreventa para ese vuelo en particular. Esta situación es la que se ha presentado en dicho vuelo. Aun cuando se ha alcanzado el cupo total de 165 reservaciones (lo cual pasa en la mayoría de los vuelos), por lo general queda un número significativo de asientos sin ocupar. Luego de recabar todos los datos posibles, el grupo de investigación de operaciones ha descubierto la razón de ello. Sólo 80% de los clientes que hacen reservaciones en este vuelo se presentan para tomarlo. El resto sacrifica el dinero que pagó (o, en la mayoría de los casos, permite que la compañía lo haga) debido a que sus planes han cambiado.

Cuando se rechaza a un cliente de este vuelo, la empresa hace los arreglos necesarios para ubicarlo en el siguiente vuelo a Chicago. Este arreglo le cuesta a la compañía 200 dólares. Además, le regala al cliente un vale por 400 dólares (sin embargo, a la compañía le cuesta sólo 300 dólares) para que lo use en un futuro vuelo. La compañía también cree conveniente que debe dársele al cliente

500 dólares adicionales por el costo intangible de la pérdida de imagen. Por lo tanto, el costo total por rechazar a un cliente es de 1 000 dólares.

El grupo de IO desea aplicar el modelo de sobreventa con el fin de determinar cuántas reservaciones deben aceptarse en este vuelo. Mediante el uso de los datos que se acaban de proporcionar, los parámetros del modelo son

$$p = 0.8, \quad r = \$300, \quad s = \$1\,000, \quad L = 150.$$

Debido a que el valor de L es muy grande, el grupo decide usar la aproximación normal de la distribución binomial. Por lo tanto, esta aproximación de n^*, el número óptimo de reservaciones a aceptar, se puede calcular mediante la ecuación

$$P\{D(n^*) \geq 150\} = \frac{r}{sp} = 0.375,$$

donde $D(n^*)$ tiene una distribución normal con una media de $\mu = np = 0.8n$ y una varianza $\sigma^2 = np(1 - p) = 0.16n$, por lo que $\sigma = 0.4\sqrt{n}$. Mediante el uso de la tabla de distribución normal proporcionada en el apéndice 5, ya que $\alpha = 0.375$ y $K_\alpha = 0.32$,

$$\frac{150 - \mu}{\sigma} = \frac{150 - 0.8n}{0.4\sqrt{n}} = 0.32,$$

lo que se reduce a

$$0.8n + 0.128\sqrt{n} - 150 = 0.$$

Despejando $\sqrt{n}$ en esta ecuación cuadrática obtenemos

$$\sqrt{n} = \frac{-0.128 + \sqrt{(0.128)^2 - 4(0.8)(-150)}}{1.6} = 13.6,$$

lo cual nos da

$$n^* = (13.6)^2 = 184.96.$$

Puesto que es necesario que x^* sea, en realidad, un entero, se redondea al entero inmediato *superior* (como lo especifica el modelo) 185.[12] La conclusión es que el número de reservaciones que se deben aceptar en este vuelo debe aumentarse de 165 a 185.

La demanda resultante $D(185)$ tendrá una media de $0.8(185) = 148$ y una desviación estándar de $0.4\sqrt{185} = 5.44$. Por ende, Transcontinental Airlines deberá poder ocupar casi o totalmente los 150 asientos del avión, sin tener que rechazar clientes, siempre que el número máximo de solicitudes de reservación sea de 185. Por lo tanto, la nueva política de aumentar el número de reservaciones aceptadas de 165 a 185 deberá aumentar de manera sustancial las ganancias que la compañía obtiene por este vuelo.

Otros modelos

Se utilizan una gran variedad de modelos en los diferentes tipos de administración de los ingresos. Con frecuencia, dichos modelos incorporan algunas de las ideas que se presentaron en los dos modelos que se mostraron en esta sección. Sin embargo, a menudo, los modelos que se usan en la práctica deben también incorporar algunas características adicionales que estos dos modelos básicos no consideran. A continuación se muestra una lista de las consideraciones prácticas que es necesario tomar en cuenta.

- El ofrecimiento de diferentes niveles de servicio (por ejemplo, cabina de primera clase, clase de negocios y clase económica en el mismo vuelo).

[12] Uno de los pasos que se deben tomar para obtener la respuesta 185 fue leer el valor de $K_\alpha = 0.32$ considerando dos decimales de la tabla de distribución normal. Sin embargo, si se utiliza la interpolación para calcular el valor de K_α con dos decimales adicionales, la solución del modelo cambiará a 186. Mediante el uso de la distribución binomial directamente en lugar de la aproximación normal se obtendrá como solución el valor 186.

- El cobro de diferentes precios por el mismo servicio (por ejemplo, descuentos a la gente mayor, niños, estudiantes, empleados, etcétera).
- El cobro de diferentes precios por el mismo servicio con base en cuánto dinero es reembolsable al cancelar con anticipación.
- Fijación de precios en forma dinámica con base en la fecha en la que se hizo la reservación y a qué tan bien se aproxima la demanda a la capacidad.
- Variación del nivel de sobreventa con base en el tiempo restante y a las cancelaciones esperadas hasta el momento en el que se ofrece el servicio.
- Tener un costo no lineal por concepto de sobreventa (por ejemplo, es posible que los primeros clientes acepten una compensación modesta por negárseles el servicio pero, a medida que pasa el tiempo, es probable que dicha compensación deba ser aumentada).
- Algunos clientes compran paquetes de servicios sometidos a diferentes términos y condiciones (por ejemplo, los clientes de una línea aérea pueden arreglar una serie de vuelos con conexión o clientes de hoteles que se hospedarán varias noches.
- Algunos clientes compran varias unidades (es decir, parejas, familias completas o grupos de personas que viajan juntas).

Representa un verdadero reto incorporar éstas y otras consideraciones prácticas en modelos más complejos. Sin embargo, ha habido un avance sorprendente por parte de investigadores y profesionales en el campo de la IO. Ésta ha sido una de las áreas más excitantes de la aplicación de esta disciplina. Adentrarnos más en este tema está fuera del alcance de este libro, pero se pueden consultar detalles en la referencia seleccionada 10, así como en las 591 referencias de este libro.

18.9 CONCLUSIONES

Los modelos de inventarios que se presentaron en el capítulo son del tipo básico, pero cumplen el propósito de mostrar la naturaleza general de estos modelos. Más aún, son representaciones suficientemente exactas de muchas situaciones reales de inventarios, por lo que, con frecuencia, resultan útiles en la práctica. En particular, los modelos EOQ se aplican de manera extensa. Algunas veces, estos modelos se han modificado para incluir algún tipo de demanda estocástica, como el modelo estocástico de revisión continua. El modelo estocástico de un solo periodo es muy conveniente para productos perecederos. Los modelos elementales de administración de los ingresos de la sección 18.8 representan un punto de partida de los tipos complejos de análisis de administración de los ingresos que se aplican en la actualidad en la industria aeronáutica y en otras industrias de servicios con características similares.

En la economía global de la actualidad, los modelos de inventarios de escalones múltiples (como los que se presentaron en la sección 18.5) juegan un papel cada vez más importante para administrar la cadena de proveedores de una compañía.

De cualquier manera, muchas situaciones de inventarios presentan complicaciones que los modelos en este capítulo no toman en cuenta, como la interacción entre los productos o tipos complicados de sistemas de inventarios con escalones múltiples. Se han formulado modelos más complejos en un intento por ajustar estas situaciones, pero todavía dejan un gran abismo entre la teoría y la práctica. El desarrollo de modelos útiles para administrar la cadena de proveedores es un área en particular activa en la investigación actual. Se ha llevado a cabo una gran cantidad de investigación acerca del desarrollo de modelos complejos de administración de los ingresos que toman en cuenta aspectos más complicados que se presentan en la práctica.

El crecimiento de hoy ocurre en el procesamiento computarizado de los datos de inventarios, junto con el crecimiento paralelo en la administración científica de inventarios.

REFERENCIAS SELECCIONADAS

1. Axsäter, S.: *Inventory Control*, 2a. ed., Springer, Nueva York, 2006.
2. Bertsimas, D. y A. Thiele: "A Robust Optimization Approach to Inventory Theory", en *Operations Research*, **54**(1): 150-168, enero-febrero de 2006.
3. Gallego, G. y O. Ozer: "A New Algorithm and a New Heuristic for Serial Supply Systems", en *Operations Research Letters*, **33**(4): 349-362, julio de 2005.

4. Harrison, T. P., H. L. Lee y J. J. Neale (eds.): *The Practice of Supply Chain Management: Where Theory and Aplication Converge*, Kluwer Academic Publishers, Boston, 2003.

5. Khouja, M: "The Single-Period (News-Vendor) Problem: Literature Review and Suggestions for Future Research", en *Omega*, **27:** 537-553, 1999.

6. Muckstadt, J. y R. Roundy: "Analysis of Multi-Stage Production Systems", pp. 59-131 en Graves, S., A. Rinnooy Kan y P. Zipken (eds.): *Handbook in Operations Research and Management Science*, vol. 4, Logistics of Production and Inventory, North-Holland, Amsterdam, 1993.

7. Sethi, S. P., H. Yan y H. Zhang: *Inventory and Supply Chain Management with Forecast Updates*, Springer, Nueva York, 2005.

8. Sherbrooke, C. C.: *Optimal Inventory Modeling of Systems: Multi-Echelon Techniques*, 2a. ed., Kluwer Academic Publishers (Springer en la actualidad), Boston, 2004.

9. Simchi-Levi, D., S. D. Wu y Z. J. Shen (eds): *Handbook of Quantitative Supply Chain Analysis*, Kluwer Academic Publishers (Springer en la actualidad), Boston, 2004.

10. Talluri, G. y K. Van Ryzin: *Theory and Practice of Yield Management*, Kluwer Academic Publishers (Springer en al actualidad), Boston, 2004.

11. Tang, C. S., C. P. Teo y K. K. Wei (eds.): *Supply Chain Analysis: A handbook on the Interaction of Information*, System and Optimization, Springer, Nueva York, 2008.

12. Tayur, S., R. Ganeshan y M. Magazine (eds.): *Quantitative Models for Supply Chain Management*, Kluwer Academic Publishers (Springer en la actualidad), Boston, 1998.

13. Tiwari, V. y S. Gavirneni: "ASP, The Art and Science of Practice: Recoupling Inventory Control Research and Practice: Guidelines for Achieving Synergy", en *Interfaces*, **37**(2): 176-186, marzo-abril de 2007.

14. Zipken, P. H.: *Foundations of Inventory Management*, McGraw-Hill, Boston, 2000.

Algunas aplicaciones de la teoría de inventarios que han ganado premios

(En el sitio en internet de este libro —www.mhhe.com/hillier— se proporciona una liga hacia estos artículos.)

A1. Arntzen, B. C., G. G. Brown, T. P. Harrison y L. L. Trafton: "Global Supply Chain Management at Digital Equipment Corporation", en *Interfaces*, **25**(1): 69-93, enero-febrero de 1995.

A2. Billington, C., G. Callioni, B. Crane, J. D. Rurak, J. U. Rapp, T. White y S. P. Willems: "Accelerating the Profitability of Hewlett-Packard's Supply Chains", en *Interfaces*, **34**(1): 59-72, enero-febrero de 1993.

A3. Flowers, A. D.: "The Modernization of Merit Brass", en *Interfaces*, **23**(1): 97-108, enero-febrero de 1993.

A4. Geraghty, M. K. y E. Johnson: "Revenue Management Saves National Car Rental", en *Interfaces*, **27**(1): 107-127, enero-febrero de 1997.

A5. Kok, T. de, F. Janssen, J. van Doremalen, E. van Wachem, M. Clerkx y W. Peeters: "Phillips Electronics Synchronizes Its Supply Chain to End the Bullwhip Effect", en *Interfaces*, **35**(1): 37-48, enero-febrero de 2005.

A6. Lin, G., M. Ettl, S. Buckley, S. Bagchi, D. D. Yao, B. L. Naccarato, R. Allan, K. Kim y L. Koenig: "Extended-Enterprise Supply-Chain Management at IBM Personal Systems Group and Other Divisions", en *Interfaces*, **30**(1): 7-25, enero-febrero de 2000.

A7. Lyon, P., R. J. Milne, R. Orzell y R. Rice: "Matching Assets with Demand in Supply-Chain Management at IBM Microelectronics", en *Interfaces*, **31**(1): 108-124, enero-febrero de 2001.

A8. Smith, B. C., J. F. Leimkuhler y R. M. Darrow: "Yield Management at American Airlines", en *Interfaces*, **22**(1): 8-31, enero-febrero de 1992.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

Worked Examples:

Ejemplos para el capítulo 18

Ejemplos de demostración incluidos en el OR Tutor:

Modelo estocástico de un solo periodo para productos perecederos, sin costo de preparación
Modelo estocástico de un solo periodo para productos perecederos, con costo de preparación

Archivos de Excel "Ch. 18 —Inventory Theory":

Plantilla del modelo básico EOQ (versión de Solver y versión analítica)
Plantilla del modelo EOQ con faltantes planeados (versión de Solver y versión analítica)
Plantilla del modelo EOQ con descuentos por cantidad (sólo versión analítica)
Plantilla del modelo estocástico de revisión continua
Plantilla del modelo estocástico de un solo periodo para productos perecederos, sin costo de instalación
Plantilla del modelo estocástico de un solo periodo para productos perecederos, con costo de instalación

Archivo de LINGO para ejemplos seleccionados

Glosario del capítulo 18

Complementos de este capítulo

Derivación de la política óptima del modelo estocástico de un solo periodo para productos perecederos, modelos estocásticos de revisión continua.

PROBLEMAS

A la izquierda de los problemas (o de sus incisos) se insertó una T cuando una de las plantillas enumeradas con anterioridad pudiera ser útil. Un asterisco en el número del problema indica que al final del libro se da, al menos, una respuesta parcial.

T **18.3-1.*** Suponga que la demanda mensual de un producto es de 30 unidades y que los artículos se retiran a una tasa constante. El costo de preparación de cada corrida de producción para reabastecer el inventario es de 15 dólares. El costo de producción es de 1 dólar por artículo y el costo de mantener un inventario es de $0.30 por artículo por mes.

a) Suponga que no se permiten faltantes; determine cada cuándo hacer las corridas de producción y su tamaño.
b) Si se permiten faltantes pero cuestan 3 dólares mensuales por artículo, determine cada cuántos días debe hacerse una corrida de producción y de qué tamaño debe ser.

T **18.3-2.** La demanda semanal de un producto es de 1 000 unidades y los artículos se retiran a una tasa constante. El costo de colocar una orden para reabastecer el inventario es de 40 dólares. El costo unitario por artículo es de 5 dólares y el costo de mantener un inventario es de $0.10 por artículo por semana.

a) Suponga que no se permiten faltantes. Determine cuándo y cuánto debe ordenarse.
b) Si se permiten faltantes semanales de 3 dólares por artículo, determine cuándo y cuánto debe ordenarse.

18.3-3.* Tim Madsen realiza las compras para Computer Center, una tienda de cómputo grande. Acaba de agregar la computadora más nueva, la Power, al inventario de la tienda. En promedio, las ventas de este modelo son de 13 unidades a la semana. Tim compra estas computadoras al fabricante a un costo unitario de 3 000 dólares y cada envío toma media semana en llegar.

Por rutina, Tim usa el modelo básico EOQ para determinar la política de inventarios que se aplicará a los productos más importantes. Estima que el costo anual de mantener los artículos es igual a 20% del costo de compra y que el costo de colocar una orden es de 75 dólares.

T ***a***) Tim usa la política de ordenar cinco Power a la vez, y cada orden se coloca a tiempo para que llegue justo cuando el inventario está por agotarse. Use la plantilla de Solver de Excel del modelo básico EOQ para determinar los costos anuales de esta política.
T ***b***) Use la misma hoja para generar una tabla que muestre cómo cambiarían estos costos si la cantidad por ordenar fuera 5, 7, 9, …, 25.
T ***c***) Utilice Solver para determinar el tamaño óptimo de la orden.
T ***d***) Use la versión analítica de la plantilla del modelo EOQ (que aplica la fórmula EOQ) para calcular la cantidad óptima. Compare estos resultados (incluya los costos) con los que obtuvo en el inciso *c*).
e) Verifique a mano su respuesta de la orden óptima del inciso *d*) mediante la aplicación de la fórmula EOQ.
f) Con base en la cantidad óptima que obtuvo, ¿cada cuándo debe ordenarse en promedio? ¿Cuál debe ser el nivel aproximado de inventario al colocarse cada orden?
g) ¿Cuánto reduce la política de inventario óptima el costo variable total anual (costo de mantener más costo administrativo de colocar órdenes) del modelo Power comparado con la política descrita en el inciso *a*)? ¿Cuál es el porcentaje de reducción?

18.3-4. La compañía de taxis Blue Cab es la principal de Maintown. Consume gasolina a una tasa constante de 10 000 galones por mes. Debido a este importante costo, la compañía tiene un convenio con Amicable Petroleum para comprar una cantidad grande de gasolina a precio de descuento de $3.50 por galón cada varios meses. El costo del convenio, que incluye colocar la gasolina en almacenamiento, es de 2 000 dólares por orden. El costo de mantener el inventario de gasolina se estima en $0.04 por galón por mes.

T ***a***) Utilice la plantilla de Excel Solver del modelo básico EOQ para determinar los costos anuales en que se incurriría si la gasolina se ordenara cada mes.
T ***b***) Emplee esta misma hoja de cálculo para generar una tabla que muestre cómo cambiarían estos costos si el número de meses entre órdenes cambiara a los siguientes valores: 1, 2, 3,…, 10.
T ***c***) Use Solver para encontrar la cantidad óptima por ordenar.
T ***d***) Utilice la versión analítica de la plantilla de Excel del modelo básico EOQ para encontrar la cantidad óptima de la orden. Compare los resultados con los del inciso *c*) (incluya los costos).
e) Verifique su respuesta de la cantidad óptima en *d*) mediante la aplicación manual de la fórmula de EOQ.

18.3-5. En el EOQ básico, use la fórmula de la raíz cuadrada para determinar cómo cambiaría Q^* con cada cambio en los costos o la

tasa de demanda. (Cada cambio es independiente si no se establece lo contrario.)

a) El costo fijo se reduce a 25% de su valor original.

b) La tasa de demanda anual se convierte en cuatro veces su valor original.

c) Ambos cambios de los incisos *a*) y *b*).

d) El costo unitario de mantener se reduce a 25% del valor original.

e) Ambos cambios de los incisos *a*) y *d*).

18.3-6.* Kris Lee, dueño y gerente de Quality Hardware Store, reevalúa su política de inventario de martillos. Debido a que vende un promedio de 50 martillos al mes, ha colocado órdenes de compra por 50 martillos con un distribuidor a un costo de 20 dólares cada uno al final de cada mes. Sin embargo, en razón de que coloca todas las órdenes de la tienda, pierde gran parte de su tiempo en esa tarea. Estima que el valor de su tiempo dedicado a ordenar martillos es de 75 dólares.

a) ¿Cuál debe ser el costo unitario de mantener martillos para que la política actual de Kris sea óptima según el modelo básico EOQ? ¿Cuál es este costo de mantener como porcentaje del costo unitario de adquisición?

T ***b***) ¿Cuál es la cantidad óptima por ordenar si el costo unitario de mantener es igual a 20% del costo unitario de adquisición? ¿Cuál es el CVT = costo variable total anual del inventario (costo de mantener más costo de ordenar)? ¿Cuál es el CVT de la política de inventarios actual?

T ***c***) Si el distribuidor entrega una orden de martillos en 5 días hábiles (de 25 promedio al mes), ¿cuál debe ser el punto de reorden (según el modelo básico EOQ)?

d) Como Kris no quiere incurrir en faltantes de artículos importantes decide agregar un inventario de seguridad de 5 martillos para protegerse de entregas tardías y ventas mayores que las usuales. ¿Cuál es su nuevo punto de reorden? ¿En cuánto se incrementa el CVT?

18.3-7.* Considere el ejemplo 1 (manufactura de bocinas para TV) de la secciones 18.1 y 18.3 que ilustran los modelos EOQ. Use EOQ con faltantes planeados para resolver este ejemplo si el costo unitario por faltante cambia a 5 dólares mensuales por bocina que falta.

T **18.3-8.** Speedy Wheels es un distribuidor de bicicletas. Su gerente de inventario, Ricky Sapolo, revisa la política de inventario de un modelo popular del que se venden 500 unidades por mes. El costo administrativo de colocar una orden al fabricante es de 1 000 dólares y el precio de compra es de 400 dólares por bicicleta. El costo de capital comprometido anual es igual a 15% del valor (basado en el precio de compra) de estas bicicletas. El costo adicional de guardar las bicicletas (incluye renta de espacio de almacén, seguros, impuestos, etc.) es de 40 dólares anuales por bicicleta.

a) Utilice el modelo EOQ básico para determinar la cantidad óptima por ordenar y el CVT de inventario al año.

b) Los clientes de Speedy Wheels (tiendas) no objetan los retrasos cortos hasta que lleguen sus órdenes. De esta forma, la administración está de acuerdo en implementar una nueva política que acepta pequeños faltantes ocasionales para reducir el costo variable total. Después de consultar con la administración, Ricky estima que el costo anual por faltantes (incluye pérdida de negocios futuros) será de 150 dólares multiplicado por el número promedio anual de bicicletas faltantes. Use el modelo EOQ con faltantes planeadas para determinar la nueva política óptima.

T **18.3-9.** Reconsidere el problema 18.3-3. Debido a la popularidad de la Power, Tim Madsen ha comprobado que los clientes están dispuestos a comprar una computadora aun cuando no se tenga en inventario siempre que se les asegure que la tendrán en un periodo razonable. Por lo tanto, ha decidido cambiar del modelo EOQ básico al EOQ con faltantes planeados, con base en un costo por faltantes anual de 200 dólares por computadora.

a) Use la plantilla de Excel Solver del modelo EOQ con faltantes planeados (con restricciones agregadas en el cuadro de diálogo de Solver de que C10:C11 = entero) para determinar la política óptima de inventarios y su CVT anual. ¿Cuál es la reducción del valor del CVT que encontró en el problema 18.3-3 (y dado al final del libro) cuando se aceptan faltantes planeados?

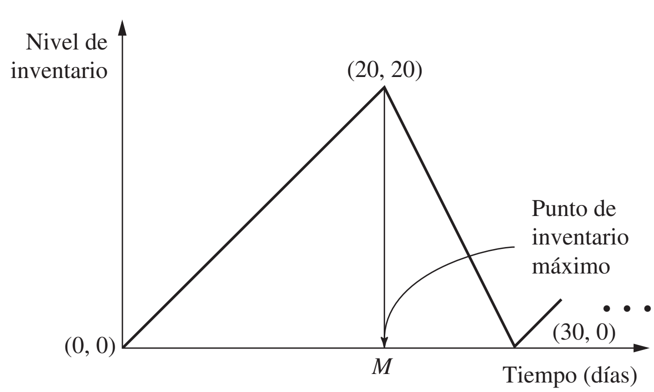
b) Utilice esta misma hoja de cálculo para generar una tabla que muestre cómo cambia el CVT y sus componentes si el faltante máximo es igual al que encontró en *a*) pero la cantidad por ordenar cambia a 15, 17, 19, …, 35.

c) Use esta misma hoja de cálculo para generar una tabla que muestre cómo cambia el CVT y sus componentes si la cantidad por ordenar es la que encontró en *a*) pero el faltante máximo cambia a 10, 12, 14,…, 30.

18.3-10. Una compañía lo contrata como consultor de IO para reevaluar la política de inventario de uno de sus productos. Ahora usan EOQ básico. Con este modelo, la cantidad óptima por ordenar es de 1 000 unidades, por lo que el nivel máximo de inventario también es de 1 000 unidades y el faltante máximo es 0.

Usted decide recomendar que cambien al modelo EOQ con faltantes planeados después de comparar el costo unitario por faltantes (p) con el costo unitario de mantener (h). Prepare una tabla para el gerente que muestre la cantidad óptima por ordenar, el nivel máximo de inventario y el faltante máximo de este modelo para las siguientes razones de p a h: $\frac{1}{3}$, 1, 2, 3, 5, 10.

18.3-11. En el EOQ básico, suponga que el reabastecimiento es uniforme (en lugar de instantáneo) a una tasa de b artículos por unidad de tiempo hasta alcanzar el tamaño del lote Q. Los artículos se retiran a una tasa de a artículos por unidad de tiempo, en donde $a < b$. Los reabastecimientos y retiros del inventario son simultáneos. Por ejemplo, si Q es 60, b es 3 al día y a es 2 al día, lo cual indica que diariamente llegan 3 unidades del día 1 al 20, 31 al 50 y así sucesivamente, mientras que las unidades se retiran a una tasa de 2 diarias todos los días. El diagrama de nivel de inventario contra el tiempo se muestra en la figura.



a) Encuentre el costo total por unidad de tiempo en términos del costo de preparación K, la cantidad producida Q, el costo unitario

c, el costo de mantener el inventario *h*, la tasa de consumo *a* y la tasa de reabastecimiento *b*.

b) Determine el tamaño del lote económico Q^*.

18.3-12.* MBI fabrica computadoras personales. Todas sus computadoras usan un disco duro que compra a Ynos. La fábrica de MBI opera 52 semanas por año y debe ensamblar 100 discos duros en las computadoras por semana. La tasa de costo de mantener de MBI es igual a 20% del valor del inventario (basado en el costo de compra). Sin importar el tamaño de la orden, el costo administrativo de colocar órdenes con Ynos se estima en 50 dólares. Y nos ofrece un descuento por órdenes grandes donde el precio de cada categoría se aplica a *todos* los lectores comprados.

Categoría de descuento	Cantidad comprada	Precio (por unidad de disco)
1	1 a 99	$100
2	100 a 499	$ 95
3	500 o más	$ 90

T ***a***) Determine la cantidad óptima por ordenar según el EOQ con descuentos por cantidad. ¿Cuál es el costo total anual que resulta?

b) Con esta cantidad por ordenar, ¿cuántas órdenes deben colocarse al año? ¿Cuál es el tiempo entre órdenes?

18.3-13. La familia Gilbreth bebe una caja de Royal Cola al día, 365 días al año. Por fortuna, un distribuidor local ofrece descuentos por cantidad en órdenes grandes, como se muestra en la tabla, donde el precio de cada categoría se aplica a *todas* las cajas compradas. El Sr. Gilbreth considera el costo de la gasolina para estimar que le cuesta alrededor de 10 dólares recoger la orden de Royal Cola. También invierte en la bolsa, donde gana un rendimiento promedio de 10% anual. Piensa que el rendimiento perdido por comprar Royal Cola en lugar de acciones es el costo de mantener su inventario.

Categoría de descuento	Cantidad comprada	Precio (por caja)
1	1 a 49	$5.00
2	50 a 99	$4.85
3	100 o más	$4.70

T ***a***) Determine la cantidad óptima por ordenar según el EOQ con descuentos por cantidad. ¿Cuál es el costo total resultante por año?

b) De acuerdo con esta cantidad por ordenar, ¿cuántas órdenes deben colocarse por año? ¿Cuál es el tiempo entre órdenes?

18.3-14. Kenicho Kaneko es gerente de un departamento de producción que usa 400 cajas de remaches al año. Para disminuir el nivel de inventario, suele ordenar sólo 50 cajas a la vez. Sin embargo, el proveedor le ofrece un descuento por órdenes de mayor cantidad de acuerdo con el siguiente programa, donde el precio de cada categoría se aplica a *todas* las cajas compradas.

Categoría de descuento	Cantidad comprada	Precio (por caja)
1	1 a 99	$8.50
2	100 a 999	$8.00
3	1 000 o más	$7.50

La compañía usa una tasa de costo anual de mantener de 20% del precio del artículo. El costo total asociado con colocar cada orden es de 80 dólares.

Kenichi ha decidido usar el modelo EOQ con descuentos por cantidad para determinar su política óptima de inventarios de remaches.

a) En cada categoría de descuento escriba una expresión del costo total anual (CT) como función de la cantidad por ordenar *Q*.

T ***b***) En cada categoría de descuento utilice la fórmula EOQ básica para calcular el valor de *Q* (factible o no) que proporciona el valor mínimo del CT. (Si lo desea, puede usar la versión analítica de la plantilla de Excel del EOQ básico para realizar este cálculo.)

c) En cada categoría de descuento use los resultados de *a*) y *b*) para determinar el valor *factible* de *Q* que proporciona el valor mínimo *factible* de CT y calcule este valor de CT.

d) Bosqueje las curvas de CT contra *Q* de cada categoría de descuento. Utilice el mismo formato que en la figura 18.3 (una línea continua si es factible y punteada si no lo es). Muestre los puntos que encontró en *b*) y *c*); no necesita más cálculos para que las curvas sean precisas en otros puntos.

e) Utilice los resultados de *c*) y *d*) para determinar la cantidad óptima por ordenar y el valor correspondiente de CT.

T ***f***) Use la plantilla de Excel de EOQ con descuentos por cantidad para verificar sus respuestas en *b*), *c*) y *e*).

g) En la categoría de descuento 2, el valor de *Q* que minimiza CT resulta ser factible. Explique por qué saber esto permite descartar la categoría 1 como candidato para proporcionar la cantidad óptima por ordenar sin hacer los cálculos que realizó en los incisos *b*) y *c*).

h) Dada la cantidad óptima por ordenar de *e*) y *f*), ¿cuántas órdenes deben colocarse al año? ¿Cuál es el tiempo entre órdenes?

18.3-15. Durante todo el año Sarah opera un kiosco de concesión en el centro. Uno de sus artículos más aceptados son los cacahuates japoneses, de los cuales vende alrededor de 200 bolsas por mes.

Sarah compra los cacahuates a Peter, 100 bolsas cada vez. Sin embargo, para alentar a los grandes compradores, Peter le ofrece descuentos por cantidades mayores con el siguiente esquema de precios, en el cual el precio de cada categoría se aplica a *todas* las bolsas compradas.

Categoría de descuento	Cantidad comprada	Precio (por bolsa)
1	1 a 199	$1.00
2	200 a 499	$0.95
3	500 o más	$0.90

Sarah desea usar el modelo EOQ con descuentos por cantidad para determinar qué cantidad debe ordenar. Para lograr su objetivo, estima una tasa de costo anual de mantener de 17% del valor de los cacahuates (con base en el precio de compra). También estima que tiene un costo fijo de 4 dólares por colocar cada orden.

Siga las instrucciones del problema 18.3-14 para analizar el problema de Sarah.

18.4-1. Suponga que debe planearse la producción para los próximos 5 meses, donde las respectivas demandas son $r_1 = 10$, $r_2 = 25$, $r_3 = 15$, $r_4 = 10$ y $r_5 = 20$. El costo de preparación es de 9 000 dólares, el costo unitario de producción es de 3 000 dólares y el costo unitario

de almacenar, 800 dólares. Determine el programa de producción óptimo que satisface las necesidades mensuales.

18.4-2. Reconsidere el ejemplo del modelo determinístico de revisión periódica de la sección 18.4. Resuelva el problema si la demanda aumenta 1 avión cada periodo.

18.4-3. Reconsidere el ejemplo del modelo determinístico de revisión periódica de la sección 18.4. Suponga que se hacen los siguientes cambios. El costo de producir cada avión varía de un periodo a otro. Además del costo de preparación de 2 millones de dólares, el costo de producir aviones en los periodos 1 o 3 es de 1.4 millones de dólares por avión, mientras que sólo cuesta 1 millón de dólares en los periodos 2 o 4.

Utilice programación dinámica para determinar cuántos aviones deben producirse (si se producen) en cada uno de los cuatro periodos para minimizar el costo total.

18.4-4.* Considere una situación en la que un producto específico se fabrica y coloca en inventario en proceso hasta que se necesita en una etapa de producción subsecuente. El número de unidades que se requieren en cada uno de los siguientes 3 meses, el costo de preparación y el costo unitario de producción en horas normales durante los próximos tres meses son los siguientes:

Mes	Requerimiento	Costo de preparación	Tiempo normal costo unitario
1	1	5	8
2	3	10	10
3	2	5	9

Por el momento hay 1 unidad en inventario y se quiere tener 2 unidades al final de los 3 meses. Se puede producir un máximo de 3 unidades en horas normales de producción cada mes, aunque se puede producir 1 unidad adicional en tiempo extra a un costo que duplica el costo de la unidad producida en tiempo normal. El costo unitario de mantener es de 2 por cada unidad extra que se almacena.

Utilice programación dinámica para determinar cuántas unidades deben producirse cada mes para minimizar el costo total.

18.5-1. Lea el artículo de referencia que describe el estudio de IO que se resume en la Viñeta de aplicación que se presentó en la sección 18.5. Describa de manera breve la forma en que se aplicó la teoría de inventarios a este estudio. Después, elabore una lista de los beneficios financieros y no financieros que arrojó dicho estudio.

18.5-2. Considere el sistema de inventario que se adapta al modelo del sistema serial de dos clases que se presentó en la sección 18.5, donde $K_1 = 25\,000$, $K_2 = 1\,500$, $h_1 = \$30$, $h_2 = \$35$ y $d = 4\,000$. Desarrolle una tabla similar a la tabla 18.1 que muestre los resultados de llevar a cabo la optimización separada de las instalaciones y la optimización simultánea de éstas. Después calcule el porcentaje de incremento del costo variable total por unidad de tiempo si los resultados que arrojó la realización de la optimización separada fueran a utilizarse en lugar de los resultados que se obtuvieron de la realización de la optimización simultánea.

18.5-3. Pronto, una compañía iniciará la fabricación de un nuevo producto. En ese momento se usará un sistema de inventarios que se ajusta al modelo de un sistema serial de dos escalones que se presentó en la sección 18.5. Actualmente existe una gran incertidumbre acerca de cuáles serán los costos de preparación y los costos de mantener inventario en las dos instalaciones, así como cuál será la tasa de demanda del nuevo producto. Por lo tanto, para comenzar a hacer planes para el nuevo sistema de inventarios es necesario verificar varias combinaciones de valores posibles de los parámetros del modelo.

Calcule Q_2^*, n^*, n y Q_1^* para las siguientes combinaciones.

***a*)** (K_1, K_2) = (\$25 000, \$1 000), (\$10 000, \$2 500) y (\$5 000, \$5 000), con $h_1 = \$25$, $h_2 = \$250$ y $d = 2\,500$.

***b*)** (h_1, h_2) = (\$10, \$500), (\$25, \$250) y (\$50, \$100), con $K_1 = \$10\,000$, $K_2 = \$2\,500$ y $d = 2\,500$.

***c*)** $d = 1\,000$, $d = 2\,500$ y $d = 5\,000$, con $K_1 = \$10\,000$ $K_2 = \$2\,500$, $h_1 = \$25$ y $h_2 = \$250$.

18.5-4. Una compañía posee una fábrica para elaborar sus productos así como una tienda para venderlos. Cierto producto nuevo se venderá de manera exclusiva en esta tienda. Su inventario del producto se reabastecerá cuando sea necesario usando el inventario de la fábrica, donde se incurre en un costo administrativo y de embarque de 200 dólares cada vez que se reabastece. La fábrica resurtirá su propio inventario del producto cuando sea necesario al prepararse para una corrida rápida de producción. Cada vez que se prepara se incurre en un costo de preparación de 5 000 dólares. El costo anual por mantener cada unidad del inventario es de 10 dólares cuando éste se conserva en la fábrica y 11 dólares cuando se mantiene en la tienda. Se espera que en la tienda se vendan 100 unidades del producto por mes. Se cumplen todos los supuestos del modelo de un sistema serial de dos escalones que se presentó en la sección 18.5 del sistema de inventarios conjunto de la fábrica y la tienda.

***a*)** Suponga que la fábrica y la tienda optimizan por separado sus propias políticas de inventarios del producto. Calcule los valores resultantes de Q_2^*, n^*, n, Q_1^* y C^*.

***b*)** Suponga que la compañía optimiza en forma simultánea la política de inventarios conjunta del producto en la fábrica y la tienda. Calcule los valores resultantes de Q_2^*, n^*, n, Q_1^* y C^*.

***c*)** Calcule el porcentaje en que decrece el costo variable total por unidad de tiempo C^* si se utiliza el enfoque descrito en el inciso *b*) en lugar del propuesto en el inciso *a*).

18.5-5. Una compañía elabora cierto producto ensamblándolo en una planta de montaje. Todos los componentes necesarios para ensamblar el producto se compran a un solo proveedor. Cada vez que la planta de montaje necesita reabastecer su inventario de componentes se recibe un embarque del proveedor con los elementos necesarios. Cada vez que se recibe un pedido, la compañía incurre en un costo de embarque de 500 dólares además del precio de compra de los componentes. Cada vez que el proveedor necesita reabastecer su propio inventario de componentes, prepara una corrida de producción rápida para fabricarlos. El costo total por la preparación de estas corridas de producción es de 50 000 dólares. El costo anual de mantener cada conjunto de componentes es de 50 dólares cuando éste es conservado por el proveedor y 60 dólares cuando lo mantiene la planta de montaje. (Es más alto en el último caso porque en esta etapa existe más capital agregado en cada conjunto de componentes.) La planta de montaje produce de manera estable 500 unidades de producto por mes. Se cumplen todos los supuestos del modelo de un sistema serial de dos escalones, el cual se describió en la sección 18.5, en el sistema de inventarios conjunto del proveedor y la planta de ensamblado.

***a*)** Suponga que el proveedor y la planta ensambladora optimizan por separado sus propias políticas de inventarios de los conjuntos de componentes. Calcule los valores resultantes de Q_2^*, n^*, n y

Q_1^*. También calcule C_1^* y C_2^*, el costo variable total por unidad de tiempo para el proveedor y la planta de montaje, respectivamente, así como $C^* = C_1^* + C_2^*$.

b) Suponga que el proveedor y la planta de ensamblaje cooperan para optimizar de manera simultánea su política de inventarios conjunta. Calcule los mismos valores que se especificaron en el inciso *a*) para esta nueva política de inventarios.

c) Compare los valores de C_1^*, C_2^* y C^* que obtuvo en los incisos *a*) y *b*). ¿Alguna de las organizaciones perdería dinero si utilizara la política de inventarios conjunta que obtuvo en el inciso *b*) en vez de las políticas separadas del inciso *a*)? De ser así, ¿qué arreglo financiero deberían hacer para inducir a la organización que pierde a firmar un contrato de suministro que aplique la política de inventarios que obtuvo en *b*)? Si se comparan los valores de C^*, ¿cuál sería el ahorro neto de las dos organizaciones si éstas pueden acordar el cumplimiento conjunto de la política óptima del inciso *b*) en lugar de las políticas óptimas separadas del inciso *a*)?

18.5-6. Considere un sistema de inventarios de tres escalones que se ajusta al modelo de un sistema serial de escalones múltiples que se presentó en la sección 18.5. Los parámetros del modelo de este sistema particular se proporcionan en seguida.

Instalación i	K_i	h_i	$d = 1\,000$
1	\$50 000	\$1	
2	\$2 000	\$2	
3	\$360	\$10	

Desarrolle una tabla como la 18.4 que muestre los resultados intermedio y final de aplicar el procedimiento de solución que se presentó en la sección 18.5 para este sistema de inventarios. Después de calcular el costo variable total por unidad de tiempo, determine el porcentaje máximo en el que este costo puede exceder al costo correspondiente de una solución óptima.

18.5-7. Siga las instrucciones del problema 18.5-5 de un modelo de inventarios de cinco escalones que se ajuste al modelo correspondiente de la sección 18.5. A continuación se presentan los parámetros del modelo.

Instalación i	K_i	h_i	$d = 1\,000$
1	\$125 000	\$2	
2	\$20 000	\$10	
3	\$6 000	\$15	
4	\$10 000	\$20	
5	\$250	\$30	

18.5-8. Reconsidere el ejemplo de un sistema de inventarios de cuatro escalones que se presentó en la sección 18.5, donde los parámetros del modelo se proporcionan en la tabla 18.2. Ahora suponga que los costos de preparación de las cuatro instalaciones de la tabla 18.2 han cambiado, y que los nuevos valores son $K_1 = \$1\,000$, $K_2 = \$5$, $K_3 = \$75$ y $K_4 = \$80$. Realice de nuevo el análisis de este ejemplo que se presentó en la sección 18.5 (como se resume en la tabla 18.4) con los nuevos costos de preparación.

18.5-9. Uno de los muchos productos elaborados por la Global Corporation se vende primordialmente en Estados Unidos. Una forma burda del producto se fabrica en una de las plantas de la corporación en Asia y después se embarca hacia una planta en Estados Unidos para el trabajo final. Luego, el producto terminado se envía al centro de distribución de la corporación en este país. El centro de distribución almacena el producto y utiliza este inventario para cumplir pedidos de varios distribuidores. Las ventas a los distribuidores permanecen relativamente uniformes a lo largo del año a una tasa de alrededor de 10 000 unidades por mes. La planta de Estados Unidos usa su inventario del producto terminado para enviar un embarque al centro de distribución cada vez que éste necesita reabastecer su inventario. El costo administrativo y de embarque asociado es de alrededor de 400 dólares por envío. Siempre que la planta de Estados Unidos necesita reabastecer su inventario, la planta asiática utiliza su inventario para enviarle un embarque. Cuando aquélla lo recibe, se prepara para una corrida de producción rápida para convertir el producto en proceso en un producto terminado. Cada vez que se repite este proceso, los costos de embarque y preparación ascienden a un total de 6 000 dólares. La planta asiática reabastece su inventario de producto en proceso cuando es necesario, preparándose para una corrida de producción rápida. Cada costo de preparación le significa una erogación de 60 000 dólares. El costo mensual de mantener cada unidad es de 3 dólares en la planta asiática, 7 dólares en la planta de Estados Unidos y 9 dólares en el centro de distribución. Se cumplen todos los supuestos del modelo del sistema serial de escalones múltiples que se presentó en la sección 18.5, para el sistema de inventarios conjunto en las tres ubicaciones del producto.

Resuelva este modelo mediante el desarrollo de una tabla como la 18.4, que muestre los resultados intermedio y final de aplicar el procedimiento de solución que se presentó en la sección 18.5. Después de calcular el costo variable total por unidad de tiempo de la solución final, determine el porcentaje máximo en el que este costo puede exceder al costo correspondiente de una solución óptima.

18.6-1. Henry Edsel es el dueño de Honest Henry, la distribuidora de autos más grande de la región. El modelo más popular es el Tritón, y los mayores costos son los asociados con los pedidos de estos autos a la fábrica y su mantenimiento en el lote. Henry pide a su gerente, Ruby Willis, que tomó un curso de IO, que use sus conocimientos para desarrollar una política de costos eficaz para saber cuándo colocar pedidos de los Tritón y cuántos pedir en cada oportunidad.

Ruby decide usar el modelo de revisión continua de la sección 18.6 para determinar la política (R, Q). Después de investigar, estima que el costo administrativo de colocar un pedido es de 1 500 dólares (papelería para órdenes), el costo de mantener por auto es de 3 000 dólares al año (15% del precio de compra de 20 000 dólares) y el costo por auto que falta es de 1 000 dólares por año (probabilidad estimada de $\frac{1}{3}$ de perder una venta y su ganancia cercana a 3 000 dólares). Después de considerar tanto la gravedad de incurrir en faltantes como el alto costo de almacenar, Ruby y Henry están de acuerdo en usar 75% de nivel de servicio (una probabilidad de 0.75 de no incurrir en faltantes durante el tiempo entre colocar un pedido y recibir los autos). Según la experiencia, también estiman que los Tritón se venden a una tasa relativamente uniforme de alrededor de 900 unidades por año.

Después de colocar un pedido, los autos se entregan en alrededor de veinte días. La mejor estimación de Ruby de la distribución de probabilidad de la demanda durante el tiempo de entrega antes de recibir el pedido es una normal con media de 50 y desviación estándar de 15.

a) Obtenga a mano la cantidad que se debe ordenar.
b) Use la tabla de la distribución normal (apéndice 5) para obtener el punto de reorden.
T *c*) Use la plantilla de Excel del OR Courseware para este modelo para verificar sus respuestas en *a*) y *b*).
d) Dadas las respuestas anteriores, ¿cuánto inventario de seguridad indica esta política?
e) Esta política puede conducir a colocar un nuevo pedido antes de la entrega de pedidos anteriores. Indique cuándo se puede presentar esta situación.

18.6-2. Uno de los artículos que más se vende en la tienda departamental J. C. Ward es un nuevo modelo de refrigerador eficiente en energía, de los cuales se venden cerca de 80 por mes. El distribuidor demora alrededor de una semana en entregar los refrigeradores a la tienda. La demanda en este tiempo tiene una distribución uniforme de entre 10 y 30. El costo de colocar una orden es de 100 dólares. El costo de mantener cada refrigerador en inventario por mes es de 15 dólares y el costo por faltante por mes se estima en 3 dólares.

El gerente de inventarios de la tienda decide usar el modelo estocástico de revisión continua de la sección 18.6, con nivel de servicio (medida 1) de 0.8, para determinar una política (R, Q).
a) Obtenga a mano R y Q.
T *b*) Utilice la plantilla de Excel correspondiente para verificar su respuesta en el inciso *a*).
c) ¿Cuál será el número promedio de faltantes por año con esta política de inventarios?

18.6-3. Con base en el modelo estocástico de revisión continua de la sección 18.6 debe tomarse una difícil decisión gerencial subjetiva sobre el nivel de servicio al cliente. El propósito de este problema es explorar las compensaciones involucradas en la toma de decisiones.

Suponga que la medida del nivel de servicio es L = probabilidad de que no ocurra un faltante durante el tiempo de entrega. Como la gerencia otorga una alta prioridad a un servicio excelente, tiende a asignar un valor alto a L, lo cual implica un inventario de seguridad grande, que contradice el deseo de eliminar el inventario innecesario. (Recuerde la *filosofía justo a tiempo* de la sección 18.3 que influye en gran medida en el pensamiento actual.) ¿Cuál es la mejor compensación entre el buen servicio y la eliminación del inventario innecesario?

Suponga que la distribución de probabilidad de la demanda en el tiempo de entrega es normal con media μ y desviación estándar σ. Entonces, el punto de reorden R es $R = \mu + K_{1-L}\sigma$, donde K_{1-L} se obtiene en el apéndice 5. La cantidad de inventario de seguridad que proporciona este punto de reorden es $K_{1-L}\sigma$. Así, si h denota el costo de mantener cada unidad en inventario, el *costo anual promedio de mantener el inventario de seguridad* (denotado por C) es $C = hK_{1-L}\sigma$.
a) Construya una tabla con cinco columnas. La primera es el nivel de servicio L, con valores: 0.5, 0.75, 0.9, 0.95, 0.99 y 0.999. Las siguientes cuatro dan C para cuatro casos. Caso 1, h = \$1 y $\sigma = 1$; caso 2, h = \$100 y $\sigma = 1$; caso 3, h = \$1 y $\sigma = 100$; caso 4, h = \$100 y $\sigma = 100$.
b) Construya otra tabla basada en la anterior con cinco renglones y las mismas cinco columnas de la primera. Cada elemento de la nueva tabla se obtiene al restar el elemento correspondiente de la primera del elemento del siguiente renglón. Por ejemplo, la primera columna de la nueva tabla es 0.75 – 0.5 = 0.25, 0.9 –0.75 = 0.15, 0.95 – 0.9 = 0.05, 0.99 – 0.95 = 0.04 y 0.999 – 0.99 = 0.009. Como estos elementos representan incrementos del nivel de servicio L, los de las siguientes cuatro columnas son los incrementos de C que se obtienen al aumentar L la cantidad que se muestra en la primera columna.
c) Según estas dos tablas, ¿qué consejo daría al gerente que debe tomar la decisión del valor de L?

18.6-4. El problema anterior describe los factores que influyen cuando se debe tomar una decisión administrativa sobre el nivel de servicio L. También señala que para cualquier valor de L, h (costo unitario anual de mantener) y σ (desviación estándar cuando la demanda durante el tiempo de entrega tiene distribución normal), el costo promedio anual de mantener el inventario de seguridad será $C = hK_{1-L}\sigma$, donde C denota el costo de mantener y K_{1-L} está dado en el apéndice 5. En consecuencia, la variabilidad de la demanda, medida por σ, tiene un efecto importante en C.

El tiempo de entrega afecta mucho el valor de σ. En particular, σ aumenta si el tiempo de entrega lo hace. El propósito de este problema es explorar esta relación.

Para concretar, suponga que el sistema de inventarios tiene los siguientes valores: $L = 0.9$, h = \$100 y $\sigma = 100$ con tiempo de entrega de 4 días; sin embargo, el proveedor que reabastece el inventario propone un cambio en el programa de entregas que afectará el tiempo de entrega. Se desea determinar cuánto cambiarán σ y C.

Se supone que en este sistema de inventarios (como es común) las demandas en días distintos son estadísticamente independientes. En este caso, la relación entre s y el tiempo de entrega está dada por la fórmula

$$\sigma = \sqrt{d}\sigma_1,$$

donde d = número de días de tiempo de entrega,
σ_1 = desviación estándar si $d = 1$.
a) Calcule C para el sistema de inventarios actual.
b) Determine σ_1. Después, calcule cuánto cambiaría C si el tiempo de entrega se redujera de 4 días a 1.
c) ¿Cuánto cambiaría C si el tiempo de entrega se duplicara de 4 a 8 días?
d) ¿Cuál debe ser el tiempo de entrega para que C se duplique de su valor actual con tiempo de entrega de 4 días?

18.6-5. Encuentre el efecto sobre el inventario de seguridad dado por el modelo estocástico de revisión continua de la sección 18.6 si se hace el siguiente cambio en el sistema de inventarios. (Los cambios son independientes.)
a) El tiempo de entrega se reduce a 0 (entrega instantánea).
b) El nivel de servicio (medida 1) disminuye.
c) El costo unitario por faltante se duplica.
d) La media de la distribución de probabilidad de la demanda durante el tiempo de entrega aumenta (sin otro cambio en la distribución).
e) La demanda durante el tiempo de entrega tiene distribución uniforme de a hacia b, pero ahora $(b - a)$ se duplica.
f) La demanda durante el tiempo de entrega tiene distribución normal con media μ y desviación estándar σ, pero ahora σ se ha duplicado.

18.6-6.* Jed Walker es el gerente de Have a Cow, un restaurante de hamburguesas en el centro. Jed compra toda la carne a Ground Chuck (un proveedor local) pero piensa cambiar a Chuck Wagon (un distribuidor nacional) porque sus precios son más bajos.

La demanda *semanal* promedio de carne es de 500 libras, con cierta variabilidad de una semana a otra. Jed estima que el costo de mantener *anual* es de 30 centavos por libra. Cuando se queda sin carne, debe comprarla en el supermercado cercano. Se estima que el alto costo de compra y la molestia que implica tienen un costo de 3 dólares por libra que le falta. Para evitar esta situación, Jed ha decidido tener suficiente inventario de seguridad para que no le falte carne antes de que llegue la entrega 95% de los ciclos de ordenar. Colocar una orden sólo requiere mandar un fax, por lo que el costo administrativo es despreciable.

El contrato de Have a Cow con Ground Chuck es el siguiente: el precio de compra es de $1.49 por libra. Se agrega un costo fijo de 25 dólares por orden, por envío. Se garantiza que llega en 2 días. Jed estima que la demanda de carne estos días tiene distribución uniforme de entre 50 y 150 libras.

Chuck Wagon propone los siguientes términos: la carne tiene un precio de $1.35 por libra. Chuck Wagon envía la carne en un camión refrigerado y cobra 200 dólares por orden más $0.10 por libra. El tiempo de envío será de alrededor de una semana, pero garantiza que no excederá de 10 días. Jed estima que la distribución de probabilidad de la demanda durante el tiempo de entrega es normal con media de 500 libras y desviación estándar de 200 libras.

T ***a***) Utilice el modelo estocástico de revisión continua que se presentó en la sección 18.6 para obtener una política (R, Q) de Have a Cow para cada proveedor alternativo.
b) Muestre cómo se calcula el punto de reorden de cada política.
c) Determine y compare el inventario de seguridad que proporcionan las dos políticas que obtuvo en *a*).
d) Determine y compare el costo anual promedio de mantener de estas dos políticas.
e) Determine y compare el costo anual de adquisición (combine precio de compra y costo de envío) de estas dos políticas.
f) Como los faltantes son poco frecuentes, los únicos costos importantes para comparar a los dos proveedores son los que obtuvo en los incisos *d*) y *e*). Agregue estos costos de cada proveedor. ¿Qué proveedor debería seleccionarse?
g) Jed desea usar la carne (que mantiene congelada) dentro del mes siguiente al que la recibe ¿Cómo influye este deseo en su elección del proveedor?

18.7-1. Lea el artículo de referencia que describe el estudio de IO que se resume en la Viñeta de aplicación que se presentó en la sección 18.7. Describa de manera breve la forma en que se aplicó la teoría de inventarios a este estudio. Después, elabore una lista de los beneficios financieros y no financieros que arrojó dicho estudio.

T **18.7-2.** Un vendedor compra periódicos a $0.55 y los vende a $0.75. El costo por faltantes es de $0.75 por periódico (ya que los compra al menudeo para satisfacer estos faltantes). El costo de mantener es de $0.01 por periódico que queda al final del día. La demanda tiene distribución uniforme entre 50 y 75. Encuentre el número óptimo de periódicos que debe comprar.

18.7-3. Freddie, el voceador, tiene un puesto. Como hay una oficina de servicios financieros cerca, uno de los periódicos que más vende es el *Financial Journal*. Cada día al amanecer compra periódicos a $1.50 y los vende a $2.50 cada uno; la siguiente mañana recibe un reembolso de $0.50 por cada periódico que no vende. El número de personas que lo compran está entre 15 y 18 por día. Freddie estima que le piden 15 periódicos 40% de los días; 16, 20% de los días; 17, 30% de los días y 18 el resto.

a) Utilice la regla de decisión de Bayes de la sección 15.2 para determinar cuál debe ser la cantidad que debe ordenar Freddie que maximiza su ganancia esperada diaria.
b) Aplique de nuevo la regla de decisión de Bayes, pero ahora con el criterio de minimizar el costo diario esperado por ordenar más o menos de lo necesario.
c) Use el modelo estocástico de un solo periodo con productos perecederos para determinar la cantidad óptima por ordenar de Freddie.
d) Con base en el modelo del inciso *c*), bosqueje la función de distribución acumulada de la demanda y muestre en la gráfica la cantidad óptima por ordenar.

18.7-4. La Donut House, de Jennifer, ofrece una gran variedad de donas. Una de ellas, una dona extragrande para compartir, está rellena de zarzamora y cubierta de chocolate. Como se requiere mucho tiempo para que la masa levante, su preparación comienza a las 4:00 a.m. por lo que debe tomarse una decisión de cuánto preparar mucho antes de saber cuántas se necesitarán. El costo de los ingredientes y mano de obra de cada dona es de 1 dólar, mientras que su precio de venta es de 3 dólares. Las que no se venden ese día se venden a un supermercado de descuento por $0.50. Durante las últimas semanas se ha registrado el número de donas vendidas cada día en 3 dólares. Los datos se resumen en la siguiente tabla.

Número vendido	Porcentaje de días
0	10%
1	15%
2	20%
3	30%
4	15%
5	10%

a) ¿Cuál es el costo unitario de ordenar menos? ¿Cuál el de ordenar más?
b) Utilice la regla de decisión de Bayes de la sección 15.2 para determinar cuántas donas deben prepararse cada día para minimizar el costo diario promedio de ordenar más o menos de lo necesario.
c) Después de graficar la distribución acumulada de la demanda, aplique el modelo estocástico de un solo periodo de bienes perecederos para determinar cuántas donas se debe preparar cada día.
d) Con base en la respuesta en *c*), ¿cuál es la probabilidad de faltantes de estas donas en un día dado?
e) Algunas familias hacen un viaje especial a Donut House para comprar una dona especial. Por lo tanto, Jennifer cree que el costo si hay faltantes puede ser mayor que la pérdida de la ganancia. En particular, puede haber un costo por la pérdida de imagen cada vez que un cliente ordena la dona y no la hay. ¿Qué tan alto tendría que ser este costo antes de que deban preparar más donas cada día de las que se determinaron en el inciso *c*)?

18.7-5.* La pastelería Swanson es famosa por producir el mejor pan de la ciudad. Tiene ventas muy altas. La demanda diaria de pan fresco tiene una distribución uniforme entre 300 y 600 panes. El pan se hornea temprano en la mañana, antes de abrir la pastelería, a un costo de 2 dólares la pieza, cada una de las cuales, ese día, se vende en 3 dólares. El pan que queda se etiqueta como pan frío y se vende con descuento por sólo $1.50 cada pieza.

a) Aplique el modelo estocástico de un periodo con bienes perecederos y calcule el nivel óptimo de servicio.

b) Aplique este modelo en forma gráfica para determinar el número óptimo de panes que se debe hornear cada mañana.

c) Debido al intervalo amplio de valores posibles en la distribución de la demanda es difícil dibujar la gráfica de *b*) con suficiente precisión para determinar un valor exacto del número óptimo de panes. Utilice álgebra para calcular ese valor exacto.

d) Dada su respuesta en *a*), ¿cuál es la probabilidad de incurrir en faltantes de pan fresco en un día dado?

e) Debido a la popularidad del pan, los clientes se decepcionan cuando ocurren faltantes. El dueño de la pastelería, Ken Swanson, concede una alta prioridad a la satisfacción de sus clientes y no le gusta tener faltantes. Cree que el análisis debe considerar también la pérdida de imagen debida a ellos. Como esta pérdida puede tener un efecto negativo en las ventas futuras, estima un costo de $1.50 por pan cada vez que un cliente no puede comprarlo porque se agotaron las existencias. Determine el número de panes que se debe hornear cada día con este cambio. ¿Cuál es la nueva probabilidad de incurrir en faltantes de pan fresco en un día dado?

18.7-6. Reconsidere el problema 18.7-5. Ahora Ken Swanson, el dueño de la pastelería, quiere realizar un análisis financiero de varias políticas de inventarios. Debe comenzar con la política que obtuvo en los primeros cuatro incisos del problema 18.7-5 (pase por alto los costos por pérdida de imagen). Como se explica en las respuestas al final del libro, esta política consiste en hornear 500 panes cada mañana, lo que da una probabilidad de $\frac{1}{3}$ de tener faltantes.

a) Calcule el ingreso por venta de pan fresco en un día en que *sí* ocurre un faltante.

b) Utilice la distribución de probabilidad de la demanda para determinar el número esperado de panes frescos vendidos los días en que *no* hay faltantes. Úselo para calcular el ingreso diario esperado por la venta de pan esos días.

c) Combine sus resultados que obtuvo en *a*) y *b*) para calcular el ingreso diario esperado por la venta de pan si se consideran todos los días.

d) Calcule el ingreso diario esperado debido al pan frío.

e) Use los resultados de *c*) y *d*) para calcular el ingreso diario esperado y después la ganancia diaria esperada (excluya los costos generales).

f) Ahora considere la política de hornear 600 panes cada mañana, de manera que nunca ocurran faltantes. Calcule la ganancia diaria esperada (excluya los costos generales) que generará esta política.

g) Considere la política de inventarios que encontró en el inciso *e*) del problema 18.7-5. Como lo implican las respuestas al final del libro, esta política es hornear 550 panes cada mañana, lo que da una probabilidad de $\frac{1}{6}$ de tener faltantes. Como esta política está a la mitad entre la considerada en los incisos *a*) hasta *e*) y la del inciso *f*), se espera que la ganancia diaria (sin costos generales ni de pérdida de imagen) también esté a la mitad entre las ganancias esperadas de esas dos políticas. Básese en este hecho para determinar la ganancia diaria esperada.

h) Ahora considere el costo de pérdida de imagen con la política que se analizó en el inciso *g*). Calcule el costo esperado de pérdida de la imagen y después la ganancia esperada si se considera este costo.

i) Repita el inciso h para la política de inventarios considerada en los incisos *a*) hasta *e*).

18.7-7. Reconsidere el problema 18.7-5. Ken Swanson, el dueño de la pastelería, desarrolló un nuevo plan para disminuir el tamaño de los faltantes. El pan se horneará dos veces al día, una antes de abrir (como antes) y la otra durante el día cuando es más fácil determinar a cuánto ascenderá la demanda ese día. La primera horneada producirá 300 panes para cubrir la demanda mínima. El tamaño de la segunda horneada se basa en una estimación de la demanda restante ese día. Se supone que esta demanda restante tiene distribución uniforme entre a y b, donde los valores de a y b se eligen cada día según las ventas. Se anticipa que $(b - a)$ casi siempre se aproximará a 75, al contrario del intervalo de 300 de la distribución de la demanda del problema 18.7-5.

a) Pase por alto los costos de pérdida de imagen [como en los incisos *a*) hasta *d*) del problema 18.7-5] y escriba la fórmula del número de panes que deben producirse en la segunda horneada en términos de a y b.

b) ¿Cuál es la probabilidad de incurrir de todas maneras en faltantes de pan fresco en un día dado? Compare esta respuesta con la probabilidad correspondiente del problema 18.7-5.

c) Cuando $b - a = 75$, ¿cuál es el tamaño máximo del faltante que puede ocurrir? ¿Cuál es el número de panes frescos que no se venderá? Compare estas respuestas con las del problema 18.7-5 donde sólo se hornea una vez al día (en la mañana temprano).

d) Ahora considere sólo el costo de ordenar menos y más. Dadas sus repuestas en *c*), compare el costo diario total esperado de ordenar más o menos con el del problema 18.7-5. ¿Qué dice esto del valor de obtener la mayor información posible sobre la demanda antes de colocar la orden final de productos perecederos?

e) Repita los incisos *a*), *b*) y *c*) pero incluya el costo de la pérdida de imagen como en el inciso *e*) del problema 18.7-5.

18.7-8. Suponga que la demanda D de una pieza de repuesto para aviones tiene distribución exponencial con media de 50, es decir,

$$\varphi_D(\xi) = \begin{cases} \dfrac{1}{50}e^{-\xi/50} & \text{para } \xi \geq 0 \\ 0 & \text{de otra manera.} \end{cases}$$

El avión caerá en la obsolescencia en un año, por lo que toda la producción se debe llevar a cabo en este momento. Por ahora, los costos de producción son de 1 000 dólares por unidad, esto es, $c = 1\,000$, pero se convierten en 10 000 dólares por unidad si se deben surtir en fecha futura, o sea, $p = 10\,000$. El costo de mantener, cargado a los excedentes al final del periodo, es de 300 dólares por unidad.

T ***a***) Determine el número óptimo de piezas de repuesto que se deben producir.

b) Suponga que el fabricante tiene 23 piezas en inventario (de un avión similar pero ya obsoleto). Determine la política óptima de inventario.

c) Suponga que no se puede determinar p, pero el fabricante desea ordenar una cantidad tal que la probabilidad de un faltante sea igual a 0.1. ¿Cuántas unidades debe ordenar?

d) Si el fabricante siguiera una política óptima que dispusiera que se debe ordenar la cantidad que se encontró en el inciso *c*), ¿cuál es el valor implícito de p?

18.7-9. Reconsidere el problema 18.6-1 de la distribuidora de Henry Edsel. El año del modelo actual está por terminar, pero la venta de los Tritón es tan buena que se agotará el inventario antes de satisfacer la demanda del año. Por fortuna, hay tiempo para colocar una orden más que se recibirá casi cuando los Tritón se acaben.

La gerente general, Ruby Willis, ahora debe decidir cuántos Tritón debe ordenar. Cada uno cuesta 20 000 dólares. Ella puede venderlos a un precio promedio de 23 000 dólares siempre que sea antes de que termine el año del modelo. Los Tritón que queden deberán venderse con descuento a 19 500 dólares. Aún más, Ruby estima que el costo de capital adicional comprometido en el inventario por un tiempo tan largo será de 500 dólares por auto, por lo cual el ingreso neto será de sólo 19 000 dólares. Como pierde 1 000 dólares en cada auto que le quede cuando cambie el modelo, concluye que no debe ordenar demasiados, pero tampoco quiere quedarse sin Tritones antes del final de año del modelo. En consecuencia, decide usar el modelo estocástico de un periodo con productos perecederos para determinar la cantidad por ordenar. Para esto, estima que el número de Tritones ordenados que se pueden vender a tiempo tiene distribución normal con media de 50 y desviación estándar de 15.

a) Determine el nivel de servicio óptimo.

b) Determine el número de Tritones que Ruby debe ordenar a la fábrica.

T **18.7-10.** Encuentre la política óptima para ordenar mediante un modelo estocástico de un periodo con costo fijo, donde la demanda tiene función de densidad de probabilidad

$$\varphi_D(\xi) = \begin{cases} \dfrac{1}{20} & \text{para } 0 \leq \xi \leq 20 \\ 0 & \text{de otra manera,} \end{cases}$$

y los costos son

$$\begin{aligned} \text{costo de mantener} &= \$1 \text{ por unidad,} \\ \text{costo por faltantes} &= \$3 \text{ por unidad,} \\ \text{costo de preparación} &= \$1.50, \\ \text{costo de producción} &= \$2 \text{ por unidad.} \end{aligned}$$

Muestre su trabajo y verifique su respuesta con la plantilla correspondiente de Excel en el OR Courseware.

T **18.7-11.** Utilice la aproximación para encontrar la política óptima del modelo estocástico de un periodo con costo fijo cuando la demanda tiene la distribución exponencial

$$\varphi_D(\xi) = \begin{cases} \dfrac{1}{25} e^{-\xi/25} & \text{para } \xi \geq 0 \\ 0 & \text{de otra manera,} \end{cases}$$

y los costos son

$$\begin{aligned} \text{costo de mantener} &= 40 \text{ centavos por unidad,} \\ \text{costo por faltante} &= \$1.50 \text{ por unidad,} \\ \text{precio de compra} &= \$1 \text{ por unidad,} \\ \text{costo de preparación} &= \$10. \end{aligned}$$

Muestre su trabajo y verifique su respuesta en la plantilla correspondiente de Excel en el OR Courseware.

18.8-1. Reconsidere el ejemplo de la línea aérea Blue Skies Airlines que se presentó en la sección 18.8. Con relación al vuelo bajo consideración, experiencias recientes indican que la demanda de la tarifa de descuento muy reducida de 200 dólares es tan alta que ésta se puede aumentar en gran medida y aún así se llenaría el avión con pasajeros de placer y de negocios. Por lo tanto, la gerencia desea conocer de qué forma cambiaría el número óptimo de asientos reservados para los clientes clase 1 si incrementara la tarifa. Realice estos cálculos para el caso de las tarifas de 300, 400, 500 y 600 dólares.

18.8-2. El crucero más popular que ofrece Luxury Cruises es uno de tres semanas en el Mediterráneo cada julio con visitas a destinos turísticos interesantes. El barco cuenta con 1 000 cabinas, por lo que es todo un reto llenarlo ya que las tarifas son elevadas. En particular, la tarifa promedio de una cabina es de 20 000 dólares, lo cual es muy elevado para un gran número de clientes potenciales. Por lo tanto, con el fin de llenar el barco, la compañía ofrece una tarifa con descuento especial para este crucero cuyo costo promedio es de 12 000 dólares por cabina en el momento en que la compañía anuncia sus cruceros con un año de antelación. La fecha límite para obtener esta tarifa con descuento es 11 meses antes del viaje, pero este descuento puede suspenderse en una fecha anterior a criterio de la compañía. A partir de esa fecha, la compañía utiliza una gran cantidad de publicidad con el fin de atraer a los clientes que buscan lujo (que son quienes hacen planes vacacionales más tarde) y están dispuestos a pagar la tarifa regular que es de alrededor de 20 000 dólares por cabina. Con base a experiencias anteriores, se estima que el número de dichos clientes que buscan lujo en este crucero tiene una distribución normal con una media de 400 y una desviación estándar de 100.

Use el modelo de tarifas de descuento con capacidad controlada que se presentó en la sección 18.8 con el fin de determinar el número máximo de cabinas que deberán venderse con la tarifa de descuento antes de reservar las cabinas restantes que se venderán a la tarifa regular.

18.8-3. Con el fin de ocupar todos los asientos en un vuelo en particular, una aerolínea ofrece una tarifa especial sin devoluciones de 100 dólares a los clientes que hagan una reservación al menos con 21 días de anticipación y cumplan con otras restricciones. A partir de esa fecha, la tarifa será de 300 dólares. Se aceptará un total de 100 reservaciones. El número de clientes que han solicitado una reservación a la tarifa completa en este vuelo en el pasado ha sido de al menos 31 y de no más de 50. Se estima que los números enteros entre 31 y 50 tienen la misma probabilidad.

Use el modelo de tarifas de descuento con capacidad controlada para determinar cuántas reservaciones deberán apartarse para los clientes que pagarían la tarifa completa.

18.8-4. Reconsidere el ejemplo de Transcontinental Airlines que se presentó en la sección 18.8. La gerencia llegó a la conclusión de que la estimación original de 500 dólares del costo intangible de la pérdida de imagen por parte del cliente rechazado es muy bajo y debe aumentarse a 1 000 dólares. Use el modelo de sobreventa para determinar el número de reservaciones que deben aceptarse en este vuelo.

18.8-5. La gerencia de Quality Airlines ha decidido basar su política de sobrerreservar en el modelo que se presentó en la sección 18.8. Es necesario que se aplique dicha política a un nuevo vuelo de Seattle a Atlanta. El avión tiene 125 asientos disponibles a una tarifa sin devolución de 250 dólares. Sin embargo, ya que comúnmente hay algunas ausencias en vuelos similares, la línea aérea debe aceptar un poco más de 125 reservaciones. En casos en los que más de 125 personas lleguen a tomar el vuelo, la línea aérea buscará voluntarios que estén dispuestos a ser asignados a un vuelo propio más tarde que tiene asientos disponibles, a cambio de ser acreedores de un certificado con un valor de 500 dólares (sin embargo, este arreglo le costará a la compañía sólo 300 dólares) en cualquier vuelo en el futuro con esta línea. La gerencia piensa que se debe sumar 300 dólares extra por el

costo intangible de la pérdida de imagen de la compañía debida a los inconvenientes generados a estos clientes.

Con base en la experiencia previa con vuelos similares que tienen alrededor de 125 reservaciones, se estima que la frecuencia relativa del número de ausencias (independientemente del número exacto de reservaciones) será la que se muestra en seguida.

Número de ausencias	Frecuencia relativa
0	0%
1	5
2	10
3	10
4	15
5	20
6	15
7	10
8	10
9	5

En lugar de usar la distribución binomial, use esta distribución directamente con el modelo de sobreventa con el fin de determinar qué cantidad de boletos de más deberá vender la compañía en este vuelo.

18.8-6. Considere el modelo de sobreventa que se presentó en la sección 18.8. Para una aplicación específica, suponga que los parámetros del modelo son $p = 0.5$, $r = \$1\ 000$, $s = \$5\ 000$ y $L = 3$. Use la distribución binomial directamente (no la aproximación normal) para calcular n^*, el número óptimo de reservaciones a aceptar, mediante prueba y error.

18.8-7. El Mountain Top Hotel es un hotel de lujo en un área popular donde se practica el ski. El hotel siempre está lleno durante los meses de invierno, por lo que las reservaciones y los pagos deben realizarse con meses de anticipación para estancias de toda una semana de sábado a sábado. Las reservaciones pueden cancelarse hasta con un mes de anticipación, esto es, no se aceptan devoluciones después de esa fecha. El hotel cuenta con 100 habitaciones y el costo de cada una de ellas por semana es de 3 000 dólares. A pesar de este elevado costo, los clientes que tienen posibilidades económicas en ocasiones dejan perder este dinero y no se presentan en el hotel debido a que modifican sus planes. En promedio, alrededor de 10% de los clientes con reservación no llegan al hotel, por lo que la gerencia ha pensado hacer una sobreventa. Sin embargo, también cree que esto debe hacerse con cuidado debido a las consecuencias tan severas que representa rechazar a un cliente que tenga una reservación. Entre dichas consecuencias se incluye el costo de hacer los arreglos con rapidez para conseguir hospedaje en un hotel de menor calidad, regalarle al cliente un vale por una estancia sin costo en el futuro y el costo intangible de una pérdida masiva de la imagen del hotel por parte del cliente furioso al que se rechazó (y seguramente comentará con sus amigos acerca de este trato tan descortés). La gerencia estima que el costo que se le deberá imponer a estas consecuencias es de 20 000 dólares.

Utilice el modelo que se presentó en la sección 18.8 incluyendo la aproximación normal de la distribución binomial para determinar qué cantidad de sobreventa debe hacer el hotel.

18.9-1. De la parte inferior de las referencias seleccionadas que se proporcionan al final del capítulo, elija una de las aplicaciones de de la teoría de inventarios ganadoras de premios. Lea este artículo y después escriba un resumen de dos páginas de la aplicación y los beneficios (incluya los no financieros) que proporciona.

18.9-2. De la parte inferior de las referencias seleccionadas que se proporcionan al final del capítulo, elija tres de las aplicaciones de la teoría de inventarios ganadoras de premios. De cada una de ellas, escriba un resumen de una página de la aplicación y los beneficios (incluya los no financieros) que proporciona.

CASOS

CASO 18.1 Actualización de control de inventarios

Robert Gates da vuelta en la esquina y sonríe al ver a su esposa arreglando los rosales de su jardín. Despacio entra al garaje, apaga el motor y llega a los brazos de su esposa.

—¿Cómo estuvo tu día? —pregunta ella.

—¡Muy bien, el negocio en la farmacia no puede estar mejor! —contesta Robert—, excepto por el tránsito de regreso. Ese tránsito puede volver loco al hombre más cuerdo. Estoy tan tenso que creo que iré adentro y me prepararé un martini relajante.

Robert entra a la casa y camina directamente a la cocina. Ve el correo en la barra y comienza a revisar algunas cuentas y anuncios hasta que ve el *OR/MS Today*. Prepara su bebida, toma la revista, va a la sala y se acomoda en su sillón reclinable. Tiene todo lo que quiere, excepto una cosa. Ve el control remoto del televisor encima del aparato. Coloca su bebida y la revista en la mesa lateral y va por el control. Ahora, con el aparato en una mano, la revista en la otra y su bebida en la mesa cerca de él, por fin domina la situación.

Enciende el televisor y cambia de canal hasta que encuentra las noticias locales. Después abre la revista y comienza a leer un artículo sobre administración científica del inventario. En ocasiones, mira el televisor para enterarse de lo último de los negocios, el clima y los deportes.

Al mismo tiempo que lee el artículo, lo distrae un comercial de cepillos dentales. Su pulso se acelera con inquietud porque el comercial de Totalee le recuerda al dentista. El comercial concluye que el cliente debe comprar un cepillo dental Totalee porque es la revolución Total…ee y es Total…mente eficaz. Y en realidad es eficaz; ¡es el cepillo dental más popular en el mercado!

En ese momento, con el artículo de inventarios y el comercial del cepillo dental en la mente, Robert tiene una idea brillante. ¡Ahora sabe cómo controlar el inventario del cepillo dental Totalee en la farmacia Nightingale!

Como gerente de inventarios, Robert ha tenido problemas para mantener suficiente inventario de *ese* cepillo dental. Ha descubierto que los clientes son muy leales a la marca Totalee, lo cual se puede deber a que tiene una patente respaldada por 9 de cada 10 dentistas. Los clientes están dispuestos a esperar

que el cepillo llegue a la farmacia Nightingale, pues los vende a un precio 20% menor que las otras farmacias locales. Esta demanda de cepillos en Nightingale significa que con mucha frecuencia se agotan. La farmacia puede recibir un envío varias horas después de hacer un pedido al almacén regional de Totalee, ya que se encuentra a una distancia de sólo 20 millas. De cualquier forma, la situación actual del inventario causa problemas porque las numerosas órdenes urgentes cuestan tiempo y papeleo innecesarios y porque los clientes se decepcionan si tienen que regresar otro día.

Robert sabe que una manera de evitar los problemas de inventarios es mediante la aplicación de la administración científica de inventarios. Toma su abrigo y las llaves del automóvil y sale rápido de la casa. Al verlo correr hacia el auto, su esposa lo llama.

—Querido, ¿a dónde vas?

—Lo siento, querida —le contesta Robert—. Acabo de descubrir una manera de controlar el inventario de un artículo crítico para la farmacia. Estoy muy emocionado porque puedo aplicar mis conocimientos de ingeniería industrial en mi trabajo. Necesito buscar los datos en la farmacia y trabajar en nuestra nueva política de inventarios. Regresaré antes de la cena.

Como el tránsito de la hora pico se había disipado, el trayecto a la farmacia casi no le toma tiempo. Abre la tienda oscura y va directamente a su oficina donde realiza una búsqueda ansiosa en los archiveros para encontrar los datos de la demanda y los costos de los cepillos Totalee durante el año anterior.

Sí. ¡Exactamente como lo sospechaba! Los datos de la demanda de cepillos dentales indican que ha sido casi constante durante meses. Ya sea invierno o verano, los clientes tienen que cepillar sus dientes y necesitan cepillos nuevos. Como el cepillo se gasta después de algunos meses de uso, los clientes siempre regresarán a comprar otro. Los datos de la demanda muestran que la farmacia Nightingale compra un promedio de 250 cepillos Totalee por mes (30 días).

Después de examinar los datos de la demanda, Robert investiga los datos de costo. Como la farmacia es tan buen cliente, Totalee le cobra su precio más bajo de sólo $1.25 por cepillo. Robert gasta alrededor de 20 minutos para colocar cada orden con Totalee. Su salario y prestaciones suman $18.75 por hora. Los costos anuales de mantener el inventario son de 12% del capital comprometido en los cepillos dentales Totalee.

***a*)** Robert decide elaborar una política de inventario que, en general, satisfaga toda la demanda, pues cree que los faltantes no valen el problema de calmar a los clientes o el riesgo de perder negocios futuros. Por lo tanto, no desea permitir faltantes planeados. Como la farmacia recibe una orden varias horas después de hacer el pedido, para simplificar, Robert supone que las entregas son instantáneas. ¿Cuál es la política óptima de inventario en estas condiciones? ¿Cuántos cepillos dentales Totalee debe ordenar Robert cada vez y con qué frecuencia? ¿Cuál es el costo variable total del inventario por año con esta política?

***b*)** Totalee ha experimentado problemas financieros porque la compañía ha perdido dinero al tratar de expandirse y fabricar otros productos de higiene personal, como cepillos para el pelo e hilo dental. Por lo tanto, la administración ha decidido cerrar el almacén que se encuentra a 20 millas de la farmacia Nightingale. Ahora la farmacia debe colocar sus pedidos con un almacén que está a 350 millas y debe esperar 6 días después de colocar la orden para recibir el envío. Dado este nuevo tiempo de entrega, ¿cuántos cepillos dentales Totalee debe ordenar Robert cada vez y cuándo debe ordenar?

***c*)** Ahora Robert se pregunta si ahorraría dinero si permite que ocurran faltantes planeados. Los clientes deberían esperar para comprar los cepillos en la farmacia Nightingale pues tienen una alta lealtad a la marca y a que la farmacia los vende más baratos. Aun cuando los clientes esperarían para comprar el cepillo Totalee en Nightingale, no quedarían muy contentos si tuviesen que regresar a la tienda para comprar el producto. Robert decide que necesita darle un valor en dólares a todas las ramificaciones negativas de los faltantes. Sabe que un empleado tendrá que calmar a cada cliente contrariado y rastrear la fecha de entrega de un nuevo envío de los cepillos Totalee. Además, cree que los clientes se molestarán debido a los inconvenientes que tendrán que enfrentar al hacer sus compras en Nightingale y quizá busquen otra tienda que les proporcione mejor servicio. Estima que los costos de manejar a los clientes disgustados y perder buena imagen y ventas futuras valen $1.50 por unidad faltante por año. Dado el tiempo de entrega de 6 días y el hecho de permitir faltantes, ¿cuántos cepillos dentales Totalee debe ordenar Robert cada vez y cuándo debe ordenarlos? ¿Cuál es el faltante máximo según esta política de inventarios? ¿Cuál es el costo variable total del inventario por año?

***d*)** Robert sabe que su estimación del costo de faltantes es sencillamente eso, una estimación. También se percata de que los empleados en ocasiones deben pasar varios minutos con cada cliente que desea comprar un cepillo dental cuando éstos están agotados. Además, ahora cree que el costo de perder buena imagen con los clientes y ventas futuras puede tener una variación amplia. Estima que el costo de tratar con clientes disgustados y perder buena imagen y ventas futuras puede variar entre 85 centavos y 25 dólares por unidad faltante por año. ¿Qué efecto tendrá cambiar la estimación del costo unitario por faltante en la política de inventarios y en el costo variable total del inventario por año que se calculó en el inciso *c*)?

***e*)** Cerrar los almacenes no ha mejorado de manera significativa la situación real de Totalee, por lo que la compañía ha decidido instituir una política de descuento para alentar las ventas. Totalee cobrará $1.25 por cepillo dental en cualquier orden de 500 cepillos o menos, $1.15 por cepillo si se ordenan más de 500 pero menos de 1 000 cepillos dentales y 1 dólar por cepillo si se ordenan 1 000 o más. Robert todavía se basa en el supuesto de un tiempo de entrega de 6 días, pero no desea que ocurran faltantes planeados. Con la nueva política de descuento, ¿cuántos cepillos dentales Totalee debe ordenar Robert cada vez y cuándo debe ordenar? ¿Cuál es el costo total del inventario (incluya los costos de compra) por año?

RESUMEN DE LOS CASOS ADICIONALES EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

CASO 18.2 Aprovechar las enseñanzas del voceador

Un joven empresario operará un puesto de petardos durante el 4 de julio. Tiene tiempo de colocar sólo una orden de los petardos que venderá en su puesto. Después de obtener los datos financieros relevantes y alguna información con la cual estimar la distribución de probabilidad de las ventas potenciales, necesita determinar cuántos paquetes de petardos debe ordenar para maximizar su ganancia esperada bajo diferentes escenarios.

CASO 18.3 Descartar el inventario excedente

American Aerospace produce motores para jet. Los faltantes frecuentes de una parte crítica han causado retrasos en la producción del motor de jet más popular, por lo que es necesario desarrollar una nueva política de inventarios para manejar esta situación. Hay un tiempo de entrega largo entre el momento en que se coloca una orden de la parte y el momento en que se recibe la cantidad ordenada. La demanda de la parte durante este tiempo de entrega es incierta, pero existen algunos datos para estimar su distribución de probabilidad. En el futuro, el nivel de inventarios de la parte estará bajo revisión continua. Ahora se debe tomar una decisión acerca del nivel de inventario en el cual se debe colocar una nueva orden y cuál debe ser la cantidad por ordenar.

CAPÍTULO 19

Procesos de decisión markovianos

El capítulo 16 se enfocó en las *cadenas de Markov* y su análisis. La mayor parte del capítulo se dedicó a las cadenas de Markov de *tiempo discreto*, es decir, a las que se observa sólo en puntos discretos del tiempo (por ejemplo, al final de cada día) y no de manera continua. Cada vez que se observa, la cadena de Markov se puede encontrar en cualquiera de varios *estados*. Dado el estado actual, una *matriz de transición* (de un paso) proporciona las probabilidades de cuál será el estado en la siguiente unidad de tiempo. Dada la matriz de transición, el capítulo 16 se centró en la *descripción del comportamiento* de una cadena de Markov, por ejemplo, encontrar las probabilidades de estado estable del estado en el que se encuentra.

Muchos sistemas importantes (como varios sistemas de colas) se pueden modelar de acuerdo con una cadena de Markov de tiempo discreto o de tiempo continuo. La descripción del comportamiento de tales sistemas, como se hizo en el capítulo 17 con los sistemas de colas, es útil para evaluar su desempeño. Sin embargo, puede ser aún más útil *diseñar la operación* del sistema para *optimizar su desempeño*, como se hizo en la sección 17.10 respecto de los sistemas de colas.

Este capítulo se enfoca precisamente en la manera en que es posible diseñar la operación de una cadena de Markov de tiempo discreto para optimizar su desempeño. Por lo tanto, en lugar de aceptar en forma pasiva el diseño de la cadena de Markov y su matriz de transición fija correspondiente, ahora se actuará. Para cada estado posible de la cadena de Markov se tomará una decisión sobre cuál de las diferentes acciones alternativas debe tomarse en ese estado. La acción elegida afecta las *probabilidades de transición* al igual que los *costos (o beneficios) inmediatos* y los *costos (o beneficios) subsecuentes* de operar el sistema. Se desea elegir las acciones óptimas para los respectivos estados al considerar tanto los costos inmediatos como los subsecuentes. Este proceso de decisión se conoce como *proceso de decisión markoviano*.

En la primera sección se presenta un ejemplo prototipo de la aplicación de un proceso de decisión markoviano. En la sección 19.2 se formula el modelo básico de estos procesos. Las siguientes tres secciones describen cómo resolverlos.

19.1 EJEMPLO PROTOTIPO

Un fabricante opera una máquina clave en el núcleo de uno de sus procesos. Debido a que se le da un uso pesado la máquina se deteriora con rapidez, lo que afecta tanto la calidad como la cantidad de producción que ella genera. Por lo tanto, al final de cada semana se realiza una inspección exhaustiva cuyo resultado es la clasificación de las condiciones de la máquina en uno de cuatro estados posibles:

Estado	Condición
0	Tan buena como nueva
1	Operable: deterioro menor
2	Operable: deterioro mayor
3	Inoperable: producción de calidad inaceptable

Después de recolectar datos históricos sobre los resultados de estas inspecciones, se hace un análisis estadístico de la evolución del estado de la máquina de un mes a otro. La siguiente matriz muestra la frecuencia relativa (probabilidad) de cada transición posible del estado en el que se encuentra en un mes (un renglón de la matriz) al estado en el que se encontrará el siguiente mes (la columna de la matriz).

Estado	**0**	**1**	**2**	**3**
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	0	$\frac{1}{2}$	$\frac{1}{2}$
3	0	0	0	1

Además, el análisis estadístico ha demostrado que estas probabilidades de transición no son afectadas por considerar también en qué estados se encontraba en meses anteriores. Esta "propiedad de falta de memoria" es la *propiedad markoviana* descrita en la sección 16.2. Por lo tanto, en el caso de la variable aleatoria X_t, que es el estado de la máquina al final del mes t, se ha concluido que el proceso estocástico $\{X_t, t = 0, 1, 2, \dots\}$ es una *cadena de Markov de tiempo discreto* cuya *matriz de transición* (de un paso) es exactamente la matriz anterior.

Como lo indica el último elemento de esta matriz de transición, una vez que la máquina deja de ser operable (entra en el estado 3), permanece inoperable. En otras palabras, el estado 3 es un *estado absorbente*. Dejar la máquina en este estado sería intolerable dado que se detendría el proceso de producción, por lo que la máquina debe ser reemplazada. (La reparación no es factible en este estado.) En consecuencia, la nueva máquina comenzaría en el estado 0.

El proceso de reemplazo toma 1 semana de manera que la producción se pierde durante este periodo. El costo de la producción perdida (ganancia perdida) es de 2 000 dólares, mientras que el costo de reemplazar la máquina es de 4 000 dólares, de manera que el costo total en el que se incurre siempre que la máquina actual entra al estado 3 es de 6 000 dólares.

Aun antes de que la máquina llegue al estado 3, puede incurrirse en costos por producir artículos defectuosos. Los costos esperados por semana por este concepto son:

Estado	**Costo esperado debido a artículos defectuosos, $**
0	0
1	1 000
2	3 000

Se han mencionado todos los costos relevantes asociados con una *política de mantenimiento*, reemplazar la máquina cuando es inoperable, pero no darle mantenimiento en otros casos. Con esta política, la evolución del estado del *sistema* (la sucesión de máquinas) todavía es una cadena de Markov, pero ahora con la matriz de transición siguiente:

Estado	**0**	**1**	**2**	**3**
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	0	$\frac{1}{2}$	$\frac{1}{2}$
3	1	0	0	0

Para evaluar esta política de mantenimiento, deben considerarse tanto los costos inmediatos en que se incurre en la semana que sigue (descritos antes), como los costos subsecuentes que resultan cuando el sistema evoluciona de este modo. Como se explicó en la sección 16.5, una medida de desempeño que se usa con amplitud para cadenas de Markov es el **costo promedio esperado por unidad de tiempo**[1] (a largo plazo).

Para calcular esta medida, primero se obtienen las *probabilidades de estado estable*, π_0, π_1, π_2 y π_3 de esta cadena de Markov con la solución del siguiente sistema de ecuaciones:

$$\pi_0 = \pi_3,$$

$$\pi_1 = \frac{7}{8}\pi_0 + \frac{3}{4}\pi_1,$$

$$\pi_2 = \frac{1}{16}\pi_0 + \frac{1}{8}\pi_1 + \frac{1}{2}\pi_2,$$

$$\pi_3 = \frac{1}{16}\pi_0 + \frac{1}{8}\pi_1 + \frac{1}{2}\pi_2,$$

$$1 = \pi_0 + \pi_1 + \pi_2 + \pi_3.$$

(Aunque este sistema de ecuaciones es lo suficientemente pequeño como para ser resuelto a mano sin gran dificultad, el procedimiento para probabilidades de estado estable en el área de cadenas de Markov del IOR Tutorial proporciona otra forma rápida de obtener esta solución.) La solución simultánea es

$$\pi_0 = \frac{2}{13}, \qquad \pi_1 = \frac{7}{13}, \qquad \pi_2 = \frac{2}{13}, \qquad \pi_3 = \frac{2}{13}.$$

De esta forma, el costo promedio esperado (a largo plazo) por semana de esta política de mantenimiento es

$$0\pi_0 + 1\,000\pi_1 + 3\,000\pi_2 + 6\,000\pi_3 = \frac{25\,000}{13} = \$1\,923.08.$$

Sin embargo, existen otras políticas de mantenimiento que deben considerarse y compararse con ésta. Por ejemplo, quizá la máquina debería reemplazarse antes de llegar al estado 3. Otra alternativa es realizar una *reparación general* a un costo de 2 000 dólares. Esta opción no es factible en el estado 3 y no mejora la máquina si está en el estado 0 o en el 1, y sólo es de interés en el estado 2. En este estado, una reparación general regresaría a la máquina al estado 1. Se requiere una semana para ello, por lo que otra consecuencia sería un gasto de 2 000 dólares que representan las ganancias perdidas al no producir.

En resumen, las decisiones posibles después de cada inspección son las siguientes:

Decisión	Acción	Estados relevantes
1	No hacer nada	0, 1, 2
2	Reparación general (el sistema regresa al estado 1)	2
3	Reemplazo (el sistema regresa al estado 0)	1, 2, 3

Para facilitar la consulta, la tabla 19.1 resume también los costos relevantes de cada decisión para cada estado en el que esa decisión puede ser de interés.

¿Cuál es la política de mantenimiento óptima? Se hará referencia a esta pregunta para ilustrar el material de las siguientes cuatro secciones.

[1] El término *a largo plazo* indica que el promedio debe interpretarse como que se tomó sobre un periodo *en extremo* largo, con lo que el efecto del estado inicial desaparece. En la sección 16.5 se analiza el hecho de que, cuando el tiempo tiende a infinito, el costo promedio *real* por unidad de tiempo en esencia converge hacia el costo promedio *esperado* por unidad de tiempo.

TABLA 19.1 Datos de costos del ejemplo prototipo

Decisión	Estado	Costo esperado por producir artículos defectuosos, $	Costo de mantenimiento, $	Costo (ganancia perdida) por producción perdida, $	Costo total por semana, $
1. No hacer nada	0	0	0	0	0
	1	1 000	0	0	1 000
	2	3 000	0	0	3 000
2. Repar. gral.	2	0	2 000	2 000	4 000
3. Reemplazar	1, 2, 3	0	4 000	2 000	6 000

19.2 MODELO DE PROCESOS DE DECISIÓN MARKOVIANOS

El modelo de los procesos de decisión markovianos considerados en este capítulo se puede resumir de la siguiente manera.

1. Se observa el estado i de una cadena de Markov de tiempo discreto después de cada transición ($i = 0, 1, \ldots, M$).
2. Después de cada observación, se selecciona una *decisión* (acción) k de un conjunto de K decisiones posibles ($k = 1, 2, \ldots, K$). (Algunas de las K decisiones pueden no ser relevantes para algunos estados.)
3. Si se elige la decisión $d_i = k$ en el estado i, se incurre en un *costo* inmediato que tiene un valor esperado C_{ik}.
4. La decisión $d_i = k$ en el estado i determina cuáles serán las *probabilidades de transición*[2] de la siguiente transición desde el estado i. Denote estas probabilidades de transición por $p_{ij}(k)$, para $j = 0, 1, \ldots, M$.
5. Una especificación de las decisiones de los estados respectivos ($d_0, d_1, \ldots, d_M$) prescribe una política para el proceso de decisión markoviano.
6. El objetivo es encontrar una *política óptima* de acuerdo con algún criterio de costo que considere tanto los costos inmediatos como los subsecuentes que resulten de la evolución futura del proceso. Un criterio común es minimizar el *costo promedio esperado por unidad de tiempo* (a largo plazo). (En la sección 19.5 se considera un criterio alternativo.)

Para relacionar esta descripción general con el ejemplo prototipo que se presentó en la sección 19.1, recuerde que la cadena de Markov que se observa ahí representa el estado (condición) de una máquina en particular. Después de cada inspección de la máquina, se elige entre tres decisiones posibles (no hacer nada, reparación general o reemplazo). El costo esperado inmediato que resulta se muestra en la columna de la derecha de la tabla 19.1 para cada combinación relevante de estados y decisiones. En la sección 19.1 se analizó una política específica $(d_0, d_1, d_2, d_3) = (1, 1, 1, 3)$, donde la decisión 1 (no hacer nada) se toma en los estados 0, 1 y 2 y la decisión 3 (reemplazo) se toma en el estado 3. Las probabilidades de transición que resultan se muestran en la última matriz de transición de la sección 19.1.

El modelo general califica como un proceso de decisión de *Markov* porque posee la propiedad markoviana que caracteriza a estos procesos. En particular, dados el estado y la decisión actuales, cualquier afirmación probabilística sobre el futuro del proceso es por completo independiente de cualquier información proporcionada sobre la historia del proceso. Esta propiedad se cumple aquí ya que 1) se trata de una cadena de Markov; 2) las nuevas probabilidades de transición dependen sólo del estado y la decisión actuales, y 3) el costo esperado inmediato también depende sólo del estado y la decisión actuales.

[2] Los procedimientos de solución dados en las siguientes dos secciones también suponen que la matriz de transición que resulta es *irreductible*.

Recuadro de aplicación

En 2003, **Bank One Corporation** fue el sexto banco más grande en Estados Unidos. Bank One Card Services, Inc., una de sus divisiones, también fue el distribuidor más grande de tarjetas Visa en ese país, de parte de Bank One y de varios miles de socios de marketing. Al año siguiente, Bank One Corporation se fusionó con *JPMorgan Chase* bajo este último nombre con el fin de formar la tercera institución bancaria más grande del país. *Chase*, a partir de ese momento, fue la palabra que se empleó como la marca de sus servicios de tarjeta de crédito.

El negocio de las tarjetas de crédito es un área de aplicación natural de IO debido a que su éxito depende directamente del equilibrio cuidadoso de varios factores cuantitativos. La tasa de porcentaje anual (APR) de cargos por concepto de intereses y la línea de crédito de las cuentas de las tarjetas influyen en su uso y en la ganancia del banco. Los clientes perciben como algo atractivo los bajos niveles de APR y las elevadas líneas de crédito. Sin embargo, los bajos niveles de APR pueden reducir la ganancia del banco, mientras que el aumento indiscriminado de las líneas de crédito incrementa la exposición del banco a la pérdida de crédito. Es un aspecto muy crítico que estos factores estén balanceados de formas diferentes para clientes distintos con base en el grado de solvencia que éstos demuestren.

Tomando estas ideas en cuenta, en 1999 la gerencia de Bank One solicitó a su grupo de IO que desarrollara el proyecto PORTICO (control y optimización del portafolio) para evaluar métodos para mejorar las ganancias de su negocio de tarjetas de crédito. El grupo de IO diseñó el sistema PORTICO usando *procesos de decisión markovianos* para seleccionar los niveles APR y las líneas de crédito de tarjetahabientes individuales que maximicen el *valor presente neto* de todo el portafolio de los clientes de tarjeta de crédito. El grupo utilizó varias variables —incluyendo el nivel de la línea de crédito, el nivel de APR y algunas variables que describen el comportamiento del cliente en cuanto a sus pagos—, para determinar el estado al cual asignar una cuenta en cualquier mes. Las probabilidades de transición se basaron en 18 meses de datos en series de tiempo en una muestra aleatoria de 3 millones de cuentas de tarjeta de crédito del portafolio del banco. Las decisiones que se deberían tomar para cada estado del proceso de decisión markoviano son el nivel APR y el nivel de la línea de crédito para esa categoría de clientes en el mes siguiente.

Después de un largo periodo de prueba se comprobó que el modelo PORTICO aumentaba en forma significativa la rentabilidad del banco. A medida que la implantación real comenzaba, se estimó que este nuevo proceso podría *incrementar las ganancias anuales en más de* **75 millones de dólares**. Esta sorprendente aplicación de los procesos de decisión markovianos le valió a Bank One ganar el prestigiado *Wagner Prize for Excellence in Operations Research Practice* en el año 2002.

Fuente: M. S. Trench, S. P. Pederson, E. T. Lau, L. Ma, H. Wang y S. K. Nair: "Managing Credit Lines and Prices for Bank One Credit Cards", e *Interfaces*, **33**(5): 4-21, septiembre-octubre de 2003. (En el sitio en internet de este libro —www.mhhe.com/hillier— se proporciona un enlace hacia este artículo.)

La descripción de una política implica dos propiedades convenientes (pero innecesarias) que en el resto del capítulo se supondrá existen (con una excepción). La primera de ellas es que una política es **estacionaria**, es decir, siempre que el sistema se encuentre en el estado i, la regla para tomar la decisión siempre es la misma sin que importe el valor del tiempo actual t. La segunda propiedad es que una política es **determinística**, esto es, siempre que el sistema se encuentre en el estado i, la regla para tomar la decisión definitivamente selecciona una decisión específica. (Debido a la naturaleza del algoritmo involucrado, la siguiente sección considera políticas *aleatorizadas*, donde se usa una distribución de probabilidad para tomar la decisión.)

De acuerdo con este marco de referencia general, ahora se regresa al ejemplo prototipo para encontrar una política óptima mediante la enumeración y comparación de todas las políticas relevantes. Al hacerlo, se denotará por R la política específica y por $d_i(R)$ la decisión correspondiente que debe tomarse en el estado i, donde las decisiones 1, 2 y 3 se describen al final de la sección anterior. Debido a que una o más de estas tres decisiones son las únicas que se considerarían en cualquier estado, los únicos valores posibles de $d_i(R)$ son 1, 2 o 3 para cualquier estado i.

Solución del ejemplo prototipo por enumeración exhaustiva

Las políticas relevantes para el ejemplo prototipo son las siguientes:

Política	Descripción verbal	$d_0(R)$	$d_1(R)$	$d_2(R)$	$d_3(R)$
R_a	Reemplazo en el estado 3	1	1	1	3
R_b	Reemplazo en el estado 3, reparación general en el estado 2	1	1	2	3
R_c	Reemplazo en los estados 2 y 3	1	1	3	3
R_d	Reemplazo en los estados 1, 2 y 3	1	3	3	3

Cada política resulta en una matriz de transición diferente, como se muestra en seguida.

	R_a			
Estado	**0**	**1**	**2**	**3**
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	0	$\frac{1}{2}$	$\frac{1}{2}$
3	1	0	0	0

	R_b			
Estado	**0**	**1**	**2**	**3**
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	1	0	0
3	1	0	0	0

	R_c			
Estado	**0**	**1**	**2**	**3**
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	1	0	0	0
3	1	0	0	0

	R_d			
Estado	**0**	**1**	**2**	**3**
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	1	0	0	0
2	1	0	0	0
3	1	0	0	0

A partir de la última columna de la tabla 19.1, los valores de C_{ik}, son los siguientes:

Decisión *k* / **Estado *i***	C_{ik} **(en miles de dólares)**		
	1	**2**	**3**
0	0	—	—
1	1	—	6
2	3	4	6
3	—	—	6

Como se indicó en la sección 16.5, se puede calcular el costo promedio esperado (a largo plazo) por unidad de tiempo, $E(C)$, a partir de la expresión

$$E(C) = \sum_{i=0}^{M} C_{ik}\pi_i,$$

donde $k = d_i(R)$ para cada i y $(\pi_0, \pi_1, \ldots, \pi_M)$ representa la distribución de estado estable de los estados del sistema según la política R que se evalúa. Después de obtener $(\pi_0, \pi_1, \ldots, \pi_M)$ según cada una de las cuatro políticas (lo que se puede hacer con el IOR Tutorial), el cálculo de $E(C)$ se resume en la siguiente tabla:

Política	$(\pi_0, \pi_1, \pi_2, \pi_3)$	$E(C)$**, en miles de dólares**
R_a	$\left(\frac{2}{13}, \frac{7}{13}, \frac{2}{13}, \frac{2}{13}\right)$	$\frac{1}{13}[2(0) + 7(1) + 2(3) + 2(6)] = \frac{25}{13} = \$1\,923$
R_b	$\left(\frac{2}{21}, \frac{5}{7}, \frac{2}{21}, \frac{2}{21}\right)$	$\frac{1}{21}[2(0) + 15(1) + 2(4) + 2(6)] = \frac{35}{21} = \$1\,667$ ← Mínimo
R_c	$\left(\frac{2}{11}, \frac{7}{11}, \frac{1}{11}, \frac{1}{11}\right)$	$\frac{1}{11}[2(0) + 7(1) + 1(6) + 1(6)] = \frac{19}{11} = \$1\,727$
R_d	$\left(\frac{1}{2}, \frac{7}{16}, \frac{1}{32}, \frac{1}{32}\right)$	$\frac{1}{32}[16(0) + 14(6) + 1(6) + 1(6)] = \frac{96}{32} = \$3\,000$

En consecuencia, la política óptima es R_b, es decir, reemplazar la máquina cuando se encuentre en el estado 3 y hacer una reparación general cuando se encuentre en el estado 2. El costo esperado (a largo plazo) por semana es de 1 667 dólares.

Si el lector desea ver **otro pequeño ejemplo**, en la sección Worked Examples del sitio en internet de este libro se proporciona uno de ellos.

En este pequeño ejemplo, es adecuado usar la enumeración exhaustiva para encontrar la política óptima, donde sólo existen cuatro políticas relevantes. Sin embargo, muchas aplicaciones tienen tantas políticas que este enfoque no es factible. Para tales casos se necesita algoritmos que puedan encontrar una política óptima de manera eficiente. En las siguientes tres secciones se consideran este tipo de algoritmos.

19.3 PROGRAMACIÓN LINEAL Y POLÍTICAS ÓPTIMAS

En la sección 19.2 se describió el tipo principal de política (llamada política *determinística*, *estacionaria*) que se usa en los procesos de decisión de Markov. Se explicó que cualquier política R se puede interpretar como una regla que prescribe la decisión $d_i(R)$ siempre que el sistema se encuentre en el estado i, para cada $i = 0, 1, \ldots, M$. Entonces R se caracteriza por los valores

$$\{d_0(R), d_1(R), \ldots, d_M(R)\}.$$

De manera equivalente, R se puede caracterizar por la asignación de valores $D_{ik} = 0$ o 1 en la matriz

$$\begin{array}{c c} & \text{Decisión } k \\ & \begin{array}{cccc} 1 & 2 & \cdots & K \end{array} \\ \text{Estado } i \begin{array}{c} 0 \\ 1 \\ \vdots \\ M \end{array} & \begin{bmatrix} D_{01} & D_{02} & \cdots & D_{0K} \\ D_{11} & D_{12} & \cdots & D_{1K} \\ \cdots & \cdots & \cdots & \cdots \\ D_{M1} & D_{M2} & \cdots & D_{MK} \end{bmatrix}, \end{array}$$

donde cada D_{ik} ($i = 0, 1, \ldots, M$ y $k = 1, 2, \ldots, K$) se define como

$$D_{ik} = \begin{cases} 1 & \text{si la decisión de } k \text{ debe tomarse en el estado } i \\ 0 & \text{de otra manera.} \end{cases}$$

Por lo tanto, cada renglón de la matriz debe contener un solo 1 y el resto de los elementos deben ser 0. Por ejemplo, la política óptima R_b del ejemplo prototipo se puede caracterizar por la matriz

$$\begin{array}{c c} & \text{Decisión } k \\ & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \text{Estado } i \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \end{array}$$

es decir, no hacer nada (decisión 1) cuando la máquina está en el estado 0 o 1, hacer una reparación general (decisión 2) en el estado 2 y reemplazar (decisión 3) en el estado 3.

Políticas aleatorizadas

La introducción de D_{ik} proporciona una motivación para formular un *modelo de programación lineal*. Se piensa que el costo esperado de una política se puede expresar como una función lineal de la D_{ik} o de alguna variable relacionada, sujeta a restricciones lineales. Desafortunadamente, los valores de D_{ik} son enteros (0 o 1) y se requieren variables continuas para la formulación de programación lineal. Este requisito se puede manejar si se amplía la interpretación de una política. La definición previa dice que se debe tomar la misma decisión cada vez que el sistema se encuentre en el estado i. La nueva interpretación de una política pedirá la determinación de una distribución de probabilidad para tomar la decisión cuando el sistema se encuentre en el estado i.

Con esta nueva interpretación, las D_{ik} ahora deben redefinirse como

$$D_{ik} = P\{\text{decisión} = k \mid \text{estado} = i\}.$$

En otras palabras, dado que el sistema está en el estado i, la variable D_{ik} es la probabilidad de elegir la decisión k como la que debe tomarse. Entonces $(D_{i1}, D_{i2}, \ldots, D_{iK})$ es la *distribución de probabilidad* de la decisión que deberá tomarse en el estado i.

Este tipo de política que usa distribuciones de probabilidad se llama **política aleatorizada**, mientras que la política que dice que $D_{ik} = 0$ o 1 recibe el nombre de *política determinística*. Las políticas aleatorizadas se pueden caracterizar de nuevo por la matriz

$$\text{Estado } i \quad \begin{matrix} \\ 0 \\ 1 \\ \vdots \\ M \end{matrix} \overset{\displaystyle\text{Decisión } k}{\begin{matrix} 1 & 2 & \cdots & K \\ \end{matrix}} \begin{bmatrix} D_{01} & D_{02} & \cdots & D_{0K} \\ D_{11} & D_{12} & \cdots & D_{1K} \\ \cdots & \cdots & \cdots & \cdots \\ D_{M1} & D_{M2} & \cdots & D_{MK} \end{bmatrix},$$

donde cada renglón suma 1, y ahora

$$0 \leq D_{ik} \leq 1.$$

A manera de ilustración, considere una política aleatorizada para el ejemplo prototipo dado por la matriz

$$\text{Estado } i \quad \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} \overset{\displaystyle\text{Decisión } k \atop 1 \quad 2 \quad 3}{\begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}}.$$

Esta política expresa que *siempre* se debe tomar la decisión 1 (no hacer nada) cuando la máquina está en el estado 0. Si se encuentra en el estado 1, se deja como está con probabilidad $\frac{1}{2}$ y se reemplaza con probabilidad $\frac{1}{2}$, de manera que puede lanzarse una moneda para elegir. Si se encuentra en el estado 2, existe una probabilidad de $\frac{1}{4}$, de que se deje como está, una probabilidad de $\frac{1}{4}$ de que se le haga una reparación general y una de $\frac{1}{2}$, de que se la reemplace. Es de suponer que se dispone de algún dispositivo aleatorio (quizá una tabla de números aleatorios) con estas probabilidades, que se pueda usar para tomar la decisión. Por último, si se encuentra que la máquina está en el estado 3, siempre se le hace una reparación general.

Al permitir políticas aleatorizadas, de forma que las D_{ik} sean variables continuas en lugar de variables discretas, se hace posible formular un modelo de programación lineal para encontrar una política óptima.

Formulación de programación lineal

Las variables de decisión convenientes (denotadas aquí por y_{ik}) para un modelo de programación lineal se definen como sigue. Para cada $i = 0, 1, \ldots, M$ y $k = 1, 2, \ldots, K$, sea y_{ik} la probabilidad incondicional de estado estable de que el sistema se encuentre en el estado i y se toma la decisión k, es decir,

$$y_{ik} = P\{\text{estado} = i \text{ y decisión} = k\}.$$

Cada y_{ik} tiene una relación cercana con la D_{ik} correspondiente puesto que, de las reglas de probabilidad condicional, se tiene

$$y_{ik} = \pi_i D_{ik},$$

donde π_i es la probabilidad de estado estable de que la cadena de Markov se encuentre en el estado i. Lo que es más

$$\pi_i = \sum_{k=1}^{K} y_{ik},$$

de manera que

$$D_{ik} = \frac{y_{ik}}{\pi_i} = \frac{y_{ik}}{\sum_{k=1}^{K} y_{ik}}.$$

Existen tres conjuntos de restricciones sobre las y_{ik}:

1. $\sum_{i=1}^{M} \pi_i = 1$ de manera que $\sum_{i=0}^{M} \sum_{k=1}^{K} y_{ik} = 1.$

2. De los resultados de las probabilidades de estado estable (vea la sección 16.5),[3]

$$\pi_j = \sum_{i=0}^{M} \pi_i p_{ij}(k)$$

de manera que

$$\sum_{k=1}^{K} y_{jk} = \sum_{i=0}^{M} \sum_{k=1}^{K} y_{ik} p_{ij}(k), \qquad \text{para } j = 0, 1, \ldots, M.$$

3. $y_{ik} \geq 0$, para $i = 0, 1, \ldots, M$ y $k = 1, 2, \ldots, K$.

El costo promedio esperado a largo plazo por unidad de tiempo es

$$E(C) = \sum_{i=0}^{M} \sum_{k=1}^{K} \pi_i C_{ik} D_{ik} = \sum_{i=0}^{M} \sum_{k=1}^{K} C_{ik} y_{ik}.$$

Entonces, el modelo de programación lineal consiste en seleccionar las y_{ik} para

$$\text{Minimizar} \qquad Z = \sum_{i=0}^{M} \sum_{k=1}^{K} C_{ik} y_{ik},$$

sujeta a las restricciones

$$(1) \qquad \sum_{i=0}^{M} \sum_{k=1}^{K} y_{ik} = 1.$$

$$(2) \qquad \sum_{k=1}^{K} y_{jk} - \sum_{i=0}^{M} \sum_{k=1}^{K} y_{ik} p_{ij}(k) = 0, \qquad \text{para } j = 0, 1, \ldots, M.$$

$$(3) \qquad y_{ik} \geq 0, \qquad \text{para } i = 0, 1, \ldots, M;\ k = 1, 2, \ldots, K.$$

En consecuencia, este modelo tiene $M + 2$ restricciones funcionales y $K(M + 1)$ variables de decisión. [En realidad, (2) proporciona una restricción *redundante*, por lo que cualquiera de estas $(M + 1)$ restricciones se puede eliminar.]

Como éste es un modelo de programación lineal, se puede resolver por el *método símplex*. Una vez que se obtienen las y_{ik}, cada una de las D_{ik} se encuentra a partir de

$$D_{ik} = \frac{y_{ik}}{\sum_{k=1}^{K} y_{ik}}.$$

[3] El argumento k se introduce en $p_{ij}(k)$ para indicar que la probabilidad de transición apropiada depende de la decisión k.

La solución óptima que se obtiene con el método símplex tiene algunas propiedades interesantes. Contendrá $M + 1$ variables básicas $y_{ik} \geq 0$. Se puede demostrar que $y_{ik} > 0$ al menos para una $k = 1, 2, \ldots, K$, para cada $i = 0, 1, \ldots, M$. De donde se sigue que $y_{ik} > 0$ para sólo una k por cada $i = 0, 1, \ldots, M$. En consecuencia, $D_{ik} = 0$ o 1.

La conclusión clave es que la política óptima que se logra con el método símplex es *determinística* y no aleatorizada. Así, cuando se permite aleatorizar las políticas no se obtienen ventajas en cuanto a mejorar la política final. Sin embargo, tiene un papel en extremo importante en esta formulación puesto que convierte variables enteras (las D_{ik}) en variables continuas para poder usar programación lineal (PL). (La analogía en *programación entera* es usar la *soltura de PL* para aplicar el método símplex y después aprovechar que se cumple la *propiedad de las soluciones enteras* para que la solución óptima de la soltura de PL sea entera de todos modos.)

Solución del ejemplo prototipo por programación lineal

Consulte el ejemplo prototipo en la sección 19.1. Las primeras dos columnas de la tabla 19.1 proporcionan las combinaciones relevantes de estados y decisiones. En consecuencia, las variables de decisión que debe incluirse en el modelo son y_{01}, y_{11}, y_{13}, y_{21}, y_{22}, y_{23} y y_{33}. (Las expresiones generales dadas para el modelo incluyen a las y_{ik} de las combinaciones *irrelevantes* de estados y decisiones, de manera que estas $y_{ik} = 0$ en una solución óptima y es lo mismo si se eliminan desde el principio.) La columna de la derecha de la tabla 19.1 proporciona los coeficientes de estas variables de la función objetivo. Las probabilidades de transición $p_{ij}(k)$ de cada combinación relevante del estado i y la decisión k también se describen en la sección 19.1.

El modelo de programación lineal que resulta es

$$\text{Minimizar} \quad Z = 1\,000y_{11} + 6\,000y_{13} + 3\,000y_{21} + 4\,000y_{22} + 6\,000y_{23} + 6\,000y_{33},$$

sujeta a

$$\begin{aligned}
y_{01} + y_{11} + y_{13} + y_{21} + y_{22} + y_{23} + y_{33} &= 1 \\
y_{01} - (y_{13} + y_{23} + y_{33}) &= 0 \\
y_{11} + y_{13} - \left(\frac{7}{8}y_{01} + \frac{3}{4}y_{11} + y_{22}\right) &= 0 \\
y_{21} + y_{22} + y_{23} - \left(\frac{1}{16}y_{01} + \frac{1}{8}y_{11} + \frac{1}{2}y_{21}\right) &= 0 \\
y_{33} - \left(\frac{1}{16}y_{01} + \frac{1}{8}y_{11} + \frac{1}{2}y_{21}\right) &= 0
\end{aligned}$$

y

$$\text{todas las } y_{ik} \geq 0.$$

Al aplicar el método símplex se obtiene la solución óptima

$$y_{01} = \frac{2}{21}, \qquad (y_{11}, y_{13}) = \left(\frac{5}{7}, 0\right), \qquad (y_{21}, y_{22}, y_{23}) = \left(0, \frac{2}{21}, 0\right), \qquad y_{33} = \frac{2}{21},$$

de manera que

$$D_{01} = 1, \qquad (D_{11}, D_{13}) = (1, 0), \qquad (D_{21}, D_{22}, D_{23}) = (0, 1, 0), \qquad D_{33} = 1.$$

Esta política indica que debe dejarse la máquina como está (decisión 1) cuando se encuentre en el estado 0 o 1, debe hacerse una reparación general (decisión 2) cuando esté en el estado 2 y debe reemplazársela (decisión 3) si está en el estado 3. Ésta es la misma política óptima que se encontró mediante la enumeración exhaustiva al final de la sección 19.2.

En la sección Worked Examples del sitio en internet de este libro se proporciona **otro ejemplo** de aplicación de la programación lineal para obtener una política óptima para un proceso de decisión markoviano.

19.4 ALGORITMO DE MEJORAMIENTO DE POLÍTICAS PARA ENCONTRAR POLÍTICAS ÓPTIMAS

Se han estudiado dos métodos para determinar una política óptima para un proceso de decisión markoviano: *enumeración exhaustiva* y *programación lineal*. La enumeración exhaustiva es útil porque es rápida y directa para manejar problemas muy pequeños. La programación lineal se puede usar para resolver problemas bastante más grandes, y existe una gran variedad de paquetes de software para aplicar el método símplex.

A continuación se presentará un tercer método llamado *algoritmo de mejoramiento de políticas*. La ventaja más importante de este método es que tiende a ser muy eficiente debido a que casi siempre llega a una solución óptima en un número relativamente pequeño de iteraciones (muchas menos que en el método símplex con una formulación de programación lineal).

Si se sigue el modelo de la sección 19.2, ocurren dos cosas como un resultado conjunto del estado actual i del sistema y de la decisión $d_i(R) = k$ cuando se opera con la política R. Se incurre en un costo (esperado) C_{ik} que depende sólo del estado observado y de la decisión que se toma. El sistema se mueve a un nuevo estado j en el siguiente periodo de observación, con una probabilidad de transición dada por $p_{ij}(k)$. Si en efecto el estado j influye en el costo en que se incurre, entonces C_{ik} se calcula de la siguiente manera. Sea

$q_{ij}(k) =$ costo esperado en el que se incurre cuando el sistema se encuentra en el estado i y se toma la decisión k y después pasa al estado j en el siguiente periodo de observación.

Entonces,

$$C_{ik} = \sum_{j=0}^{M} q_{ij}(k)p_{ij}(k).$$

Preliminares

Con referencia a la descripción y notación de los procesos markovianos de decisión que se explicaron al principio de la sección 19.2, se puede demostrar que, para cualquier política dada R, existen valores $g(R)$, $v_0(R)$, $v_1(R)$, …, $v_M(R)$ que satisfacen

$$g(R) + v_i(R) = C_{ik} + \sum_{j=0}^{M} p_{ij}(k)\, v_j(R), \qquad \text{para } i = 0, 1, 2, \ldots, M.$$

Se dará ahora una justificación heurística de estas relaciones y una interpretación de estos valores.

Denote por $v_i^n(R)$ el costo total esperado de un sistema que se inicia en el estado i (en el primer periodo de observación) y opera durante n periodos. Entonces, $v_i^n(R)$ consta de dos componentes: C_{ik}, el costo en el que se incurre durante el primer periodo de observación y $\sum_{j=0}^{M} p_{ij}(k)\, v_j^{n-1}(R)$, el costo total esperado del sistema al operar los $n - 1$ periodos restantes. De aquí se obtiene la *ecuación recursiva*

$$v_i^n(R) = C_{ik} + \sum_{j=0}^{M} p_{ij}(k)\, v_j^{n-1}(R), \qquad \text{para } i = 0, 1, 2, \ldots, M,$$

donde $v_i^1(R) = C_{ik}$ para toda i.

Será útil explorar el comportamiento de $v_i^n(R)$ a medida que n crece. Recuerde que el costo promedio esperado (a largo plazo) por unidad de tiempo al seguir cualquier política R se puede expresar como

$$g(R) = \sum_{i=0}^{M} \pi_i C_{ik},$$

que es independiente del estado inicial i. Entonces, $v_i^n(R)$ se comporta aproximadamente como $n\, g(R)$ para n grande. En realidad, si se pasan por alto las pequeñas fluctuaciones, $v_i^n(R)$ se puede expresar como la suma de dos componentes

$$v_i^n(R) \approx n\, g(R) + v_i(R),$$

donde la primera componente es independiente del estado inicial y la segunda depende de ese estado. Entonces, $v_i(R)$ se puede interpretar como el efecto sobre el costo total esperado debido a que el proceso se inicia en el estado i. En consecuencia,

$$v_i^n(R) - v_j^n(R) \approx v_i(R) - v_j(R),$$

de manera que $v_i(R) - v_j(R)$ es una medida del efecto de comenzar en el estado i en vez de hacerlo en el estado j.

Cuando n crece lo suficiente, se puede sustituir $v_i^n(R) = n\ g(R) + v_i(R)$ y $v_j^{n-1}(R) = (n - 1)g(R) + v_j(R)$ en la *ecuación recursiva*. Esta operación conduce al sistema de ecuaciones dado en el primer párrafo de esta subsección.

Observe que este sistema tiene $M + 1$ ecuaciones con $M + 2$ incógnitas, por lo que se puede seleccionar una de estas variables de manera arbitraria. Por convención, se elegirá $v_M(R)$ igual a cero. Por lo tanto, al resolver el sistema de ecuaciones lineales se puede obtener $g(R)$, el costo promedio esperado a largo plazo por unidad de tiempo si se sigue la política R. En principio, se puede enumerar todas las políticas y encontrar aquella que minimiza $g(R)$. Sin embargo, aun para un número moderado de estados y decisiones, esta técnica es tediosa. Por fortuna, existe un algoritmo que se puede usar para evaluar las políticas y encontrar la óptima sin tener que hacer la enumeración completa, que se describe a continuación.

Algoritmo de mejoramiento de la política

El primer paso del algoritmo es elegir una política arbitraria R_1. Después se resuelve el sistema de ecuaciones para encontrar los valores $g(R_1)$, $v_0(R)$, $v_1(R)$, …, $v_{M-1}(R)$ [con $v_M(R) = 0$]. Este paso se llama *determinación del valor*. Se construye una política mejor denotada por R_2. Este paso se llama *mejoramiento de la política*. Los dos pasos anteriores constituyen una iteración del algoritmo. Con la nueva política R_2 se realiza otra iteración. Este proceso continúa hasta que dos iteraciones sucesivas lleven a políticas idénticas, con lo que se obtiene la óptima. En seguida se describen los detalles.

Resumen del algoritmo de mejoramiento de la política

Paso inicial: se elige una política inicial de prueba R_1. Se hace $n = 1$.
Iteración n:
Paso 1: Determinación del valor: para la política R_n se usa $p_{ij}(k)$, C_{ik} y $v_M(R_n) = 0$ para resolver el sistema de $M + 1$ ecuaciones

$$g(R_n) = C_{ik} + \sum_{j=0}^{M} p_{ij}(k)\ v_j(R_n) - v_i(R_n), \qquad \text{para } i = 0, 1, \ldots, M,$$

para todos los $M + 1$ valores desconocidos de $g(R_n)$, $v_0(R_n)$, $v_1(R_n)$, . . . , $v_{M-1}(R_n)$.
Paso 2: Mejoramiento de la política: con los valores actuales de $v_i(R_n)$ calculados para la política R_n, se encuentra la política alternativa R_{n+1} tal que, para cada estado i, $d_i(R_{n+1}) = k$ es la decisión que minimiza

$$C_{ik} + \sum_{j=0}^{M} p_{ij}(k)\ v_j(R_n) - v_i(R_n),$$

es decir, para *cada* estado i se quiere

$$\underset{k=1,\,2,\,\ldots\,,\,k}{\text{Minimizar}} \qquad \left[C_{ik} + \sum_{j=0}^{M} p_{ij}(k)\ v_j(R_n) - v_i(R_n)\right],$$

y después se establece $d_i(R_{n+1})$ igual al valor de k que proporciona el mínimo. Este procedimiento define una nueva política R_{n+1}.
Prueba de optimalidad: la política actual R_{n+1} es óptima si es idéntica a la política R_n. Si lo es, el algoritmo se detiene. De otra manera, se establece $n = n + 1$ y se realiza otra iteración.

Dos propiedades clave de este algoritmo son:

1. $g(R_{n+1}) \leq g(R_n)$, para $n = 1, 2, \ldots$
2. El algoritmo termina con una solución óptima en un número finito de iteraciones.[4]

Solución del ejemplo prototipo mediante el algoritmo de mejoramiento de la política

Consulte el ejemplo prototipo que se presentó en la sección 19.1; a continuación se describe la aplicación del algoritmo.

Paso inicial. Para la política inicial de prueba se elige, de manera arbitraria, la política que dice que se reemplace la máquina (decisión 3) cuando se encuentra en el estado 3, pero que no se haga nada (decisión 1) en otros estados. La matriz de transición y los costos de esta política se resumen como sigue.

Política R_1

Estado	Decisión
0	1
1	1
2	1
3	3

Matriz de transición

Estado	0	1	2	3
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	0	$\frac{1}{2}$	$\frac{1}{2}$
3	1	0	0	0

Costos

Estado	C_{ik}
0	0
1	1 000
2	3 000
3	6 000

Iteración 1. Con esta política, el paso de determinación del valor requiere resolver el siguiente sistema de cuatro ecuaciones simultáneas para $g(R_1)$, $v_0(R_1)$, $v_0(R_1)$ y $v_2(R_1)$ [con $v_3(R_1) = 0$].

$$g(R_1) = \qquad + \frac{7}{8}v_1(R_1) + \frac{1}{16}v_2(R_1) - v_0(R_1).$$

$$g(R_1) = 1\,000 \qquad + \frac{3}{4}v_1(R_1) + \frac{1}{8}v_2(R_1) - v_1(R_1).$$

$$g(R_1) = 3\,000 \qquad + \frac{1}{2}v_2(R_1) - v_2(R_1).$$

$$g(R_1) = 6\,000 + v_0(R_1).$$

La solución simultánea de este sistema de ecuaciones es

$$g(R_1) = \frac{25\,000}{13} = 1\,923$$

$$v_0(R_1) = -\frac{53\,000}{13} = -4\,077$$

$$v_1(R_1) = -\frac{34\,000}{13} = -2\,615$$

$$v_2(R_1) = \frac{28\,000}{13} = 2\,154.$$

[4] Esta terminación está garantizada con los supuestos del modelo que se presentó en la sección 19.2, entre los cuales se incluye, en particular, el supuesto (implícito) de un número finito de estados ($M + 1$) y un número finito de decisiones (K), pero no necesariamente para modelos más generales. Vea R. Howard, *Dynamic Programming and Markov Processes*, M.I.T. Press, Cambridge, MA., 1960. Vea también pp. 1291-1293 en A. F. Veinott, Jr., "On Finding Optimal Policies in Discrete Dynamic Programming with No Discounting", en *Annals of Mathematical Statistics*, **37**: 1284-1294, 1966.

Ahora se puede aplicar el paso 2 (mejoramiento de la política). Es necesario encontrar una política mejorada R_2, tal que la decisión k en el estado i minimice la expresión correspondiente:

Estado 0: $C_{0k} - p_{00}(k)(4\ 077) - p_{01}(k)(2\ 615) + p_{02}(k)(2\ 154) + 4\ 077$
Estado 1: $C_{1k} - p_{10}(k)(4\ 077) - p_{11}(k)(2\ 615) + p_{12}(k)(2\ 154) + 2\ 615$
Estado 2: $C_{2k} - p_{20}(k)(4\ 077) - p_{21}(k)(2\ 615) + p_{22}(k)(2\ 154) - 2\ 154$
Estado 3: $C_{3k} - p_{30}(k)(4\ 077) - p_{31}(k)(2\ 615) + p_{32}(k)(2\ 154)$.

En realidad, en el estado 0, la única decisión permitida es la decisión 1 (no hacer nada), por lo que no se necesita hacer cálculos. De manera similar, se sabe que la decisión 3 (reemplazar) debe tomarse en el estado 3. En consecuencia, sólo los estados 1 y 2 requieren el cálculo de los valores de estas expresiones de las diferentes decisiones.

En el caso del estado 1, las decisiones posibles son 1 y 3. Para cada una se muestran los valores de C_{1k}, $p_{1j}(k)$ y el valor resultante de la expresión.

Decisión	Estado 1					
	C_{1k}	$p_{10}(k)$	$p_{11}(k)$	$p_{12}(k)$	$p_{13}(k)$	Valor de la expresión
1	1 000	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	1 923 ← Mínimo
3	6 000	1	0	0	0	4 538

Como la decisión 1 minimiza la expresión, se elige como la decisión que debe tomarse en el estado 1 para la política R_2 (igual que para la política R_1).

Los resultados correspondientes para el estado 2 se muestran en seguida para las tres posibles decisiones.

Decisión	Estado 2					
	C_{2k}	$p_{20}(k)$	$p_{21}(k)$	$p_{22}(k)$	$p_{23}(k)$	Valor de la expresión
1	3 000	0	0	$\frac{1}{2}$	$\frac{1}{2}$	1 923
2	4 000	0	1	0	0	−769 ← Mínimo
3	6 000	1	0	0	0	−231

Por lo tanto, se elige la decisión 2 como la que se debe tomar en el estado 2 para la política R_2. Observe que esto es diferente de la política R_1.

A continuación se resumen los resultados de la nueva política, su matriz de transición y sus costos.

Política R_2

Estado	Decisión
0	1
1	1
2	2
3	3

Matriz de transición

Estado	0	1	2	3
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	1	0	0
3	1	0	0	0

Costos

Estado	C_{ik}
0	0
1	1 000
2	4 000
3	6 000

Como esta política no es idéntica a la política R_1, la prueba de optimalidad señala que se debe realizar otra iteración.

Iteración 2. En el paso 1 (determinación del valor), las ecuaciones a resolver para esta política son las siguientes.

$$g(R_2) = \qquad + \frac{7}{8}v_1(R_2) + \frac{1}{16}v_2(R_2) - v_0(R_2).$$

$$g(R_2) = 1\,000 \quad + \frac{3}{4}v_1(R_2) + \frac{1}{8}v_2(R_2) - v_1(R_2).$$

$$g(R_2) = 4\,000 \quad + \quad v_1(R_2) \qquad - v_2(R_2).$$

$$g(R_2) = 6\,000 + v_0(R_2).$$

La solución simultánea de este sistema de ecuaciones es

$$g(R_2) = \frac{5\,000}{3} = 1\,667$$

$$v_0(R_2) = -\frac{13\,000}{3} = -4\,333$$

$$v_1(R_2) = -3\,000$$

$$v_2(R_2) = -\frac{2\,000}{3} = -667.$$

Ahora se puede aplicar el paso 2 (mejorar la política). En los casos de los dos estados con más de una decisión posible, las expresiones que se quieren minimizar son

Estado 1: $C_{1k} - p_{10}(k)(4\,333) - p_{11}(k)(3\,000) - p_{12}(k)(667) + 3\,000$

Estado 2: $C_{2k} - p_{20}(k)(4\,333) - p_{21}(k)(3\,000) - p_{22}(k)(667) + 667.$

La primera iteración proporciona la mayor parte de los datos necesarios (probabilidades de transición y C_{ik}) para determinar la nueva política, excepto por los valores de cada una de estas expresiones para cada decisión posible. Estos valores son

Decisión	Valor del estado 1	Valor del estado 2
1	1 667	3 333
2	—	1 667
3	4 667	2 334

Como la decisión 1 minimiza la expresión para el estado 1 y la decisión 2 minimiza la expresión para el estado 2, la siguiente política de prueba R_3 es

Política R_3

Estado	Decisión
0	1
1	1
2	2
3	3

Observe que la política R_3 es idéntica a la política R_2. En este caso, la prueba de optimalidad indica que esta política es óptima y el algoritmo termina.

En el OR Tutor se incluye **otro ejemplo** que ilustra la aplicación de este algoritmo. Asimismo, en la sección Worked Examples del sitio en internet de este libro se puede encontrar **un ejemplo adicional**. El IOR Tutorial también incluye una rutina *interactiva* para facilitar el aprendizaje y la aplicación del algoritmo.

19.5 CRITERIO DEL COSTO DESCONTADO

En todo este capítulo se han medido las políticas según su costo promedio esperado (a largo plazo) por unidad de tiempo. Ahora se estudiará una medida alternativa de desempeño: el **costo descontado total esperado**.

Como se explicó en la sección 18.2, esta medida usa un *factor de descuento* α, donde $0 < \alpha < 1$. Este factor de descuento α se puede interpretar como igual a $1/(1 + i)$, donde i es la tasa de interés actual por periodo. En consecuencia, α es el *valor presente* de una unidad de costo de un periodo en el futuro. En forma similar, α^m es el *valor presente* de una unidad de costo dentro de m periodos en el futuro.

Este *criterio de costo descontado* es preferible al *criterio de costo promedio* cuando los periodos de la cadena de Markov son suficientemente largos como para que el *valor del dinero en el tiempo* se deba tomar en cuenta cuando los costos en periodos futuros se suman al costo en el periodo actual. Otra ventaja es que el criterio de costo descontado se puede adaptar con facilidad para manejar procesos de decisión markovianos de *periodo finito* donde la cadena de Markov deja de operar después de cierto número de periodos.

Tanto la técnica de mejoramiento de la política como el enfoque de programación lineal se pueden aplicar aquí con algunos ajustes relativamente menores al caso del costo promedio, como se describirá en seguida. Después se presentará otra técnica, llamada *método de aproximaciones sucesivas*, para aproximarse con rapidez a una política óptima.

Algoritmo de mejoramiento de la política

Para derivar las expresiones necesarias de los pasos de determinación del valor y mejoramiento de una política se adopta el punto de vista de *programación dinámica probabilística* (descrita en la sección 10.4). En particular, para cada estado i ($i = 0, 1, \ldots, M$) de un proceso de decisión markoviano que opera bajo la política R, sea $V_i^n(R)$ el *costo descontado total esperado* cuando el proceso se inicia en el estado i (al inicio del primer periodo de observación) y opera durante n periodos. Entonces, $V_i^n(R)$ tiene dos componentes: C_{ik}, el costo en el que se incurre durante el primer periodo de observación, y $\alpha \sum_{j=0}^{M} p_{ij}(k) V_j^{n-1}(R)$, el costo total descontado esperado del proceso al operar los $n - 1$ periodos restantes. Para cada $i = 0, 1, \ldots, M$, esto conduce a la ecuación recursiva

$$V_i^n(R) = C_{ik} + \alpha \sum_{j=0}^{M} p_{ij}(k) V_j^{n-1}(R),$$

con $V_i^1(R) = C_{ik}$, que se parece mucho a las relaciones recursivas de programación dinámica probabilística que se encontraron en la sección 10.4.

Conforme n se aproxima a infinito, esta relación recursiva converge a

$$V_i(R) = C_{ik} + \alpha \sum_{j=0}^{M} p_{ij}(k) V_j(R), \qquad \text{para } i = 0, 1, \ldots, M,$$

en la cual $V_i(R)$ se puede interpretar como el costo descontado total esperado cuando el proceso comienza en el estado i y continúa en operación de manera indefinida. Se tienen $M + 1$ ecuaciones y $M + 1$ incógnitas, de manera que la solución simultánea de este sistema de ecuaciones proporciona el valor de $V_i(R)$.

Con propósitos de ilustración, considere de nuevo el ejemplo prototipo de la sección 19.1. Según el criterio del costo promedio, en las secciones 19.2, 19.3 y 19.4 se determinó que la política óptima es no hacer nada si el proceso se encuentra en los estados 0 y 1, hacer una reparación general en el estado 2 y reemplazar en el estado 3. De acuerdo con el criterio de costo descontado, con $\alpha = 0.9$, esta política resulta en el siguiente sistema de ecuaciones:

$$V_0(R) = \qquad + 0.9\left[\quad \frac{7}{8}V_1(R) + \frac{1}{16}V_2(R) + \frac{1}{16}V_3(R)\right]$$

$$V_1(R) = 1\,000 + 0.9\left[\quad \frac{3}{4}V_1(R) + \frac{1}{8}V_2(R) + \frac{1}{8}V_3(R)\right]$$

$$V_2(R) = 4\,000 + 0.9[\qquad\qquad V_1(R)]$$
$$V_3(R) = 6\,000 + 0.9[V_0(R)].$$

La solución simultánea es

$$V_0(R) = 14\,949$$
$$V_1(R) = 16\,262$$
$$V_2(R) = 18\,636$$
$$V_3(R) = 19\,454.$$

Entonces, si se supone que el sistema comienza en el estado 0, el costo descontado total esperado es de 14 949 dólares.

Este sistema de ecuaciones proporciona las expresiones necesarias para el algoritmo de mejoramiento de la política. Después de resumir este algoritmo en términos generales, se usará para verificar si esta política, en particular, todavía es óptima bajo el criterio del costo descontado.

Resumen del algoritmo de mejoramiento de la política (criterio del costo descontado)

Paso inicial: Se elige una política de prueba inicial arbitraria R_1. Se establece $n = 1$.
Iteración n:
Paso 1: Determinación del valor: Para la política R_n, se utilizan $p_{ij}(k)$ y C_{ik} para resolver el sistema de $M + 1$ ecuaciones

$$V_i(R_n) = C_{ik} + \alpha \sum_{j=0}^{M} p_{ij}(k)V_j(R_n), \qquad \text{para } i = 0, 1, \ldots, M,$$

para todos los $M + 1$ valores desconocidos de $V_0(R_n), V_1(R_n), \ldots, V_M(R_n)$.
Paso 2: Mejoramiento de la política: con los valores actuales de $V_i(R_n)$, se encuentra la política alternativa R_{n+1} tal que, para cada estado i, $d_i(R_{n+1}) = k$ es la decisión que minimiza

$$C_{ik} + \alpha \sum_{j=0}^{M} p_{ij}(k)V_j(R_n),$$

esto es, para *cada* estado i, se debe

$$\underset{k=1,\,2,\,\ldots,\,K}{\text{Minimizar}} \quad \left[C_{ik} + \alpha \sum_{j=0}^{M} p_{ij}(k)V_j(R_n)\right],$$

y después se establece $d_i(R_{n+1})$ igual al valor de k que minimiza. Este procedimiento define una nueva política óptima R_{n+1}.
Prueba de optimalidad: la política actual R_{n+1} es óptima si es idéntica a la política R_n. Si lo es, el algoritmo se detiene. De otra manera se establece $n = n + 1$ y se realiza otra iteración.

Las tres propiedades clave de este algoritmo son las siguientes:

1. $V_i(R_{n+1}) \leq V_i(R_n)$, para $i = 0, 1, \ldots, M$ y $n = 1, 2, \ldots$
2. El algoritmo termina con una política óptima en un número finito de iteraciones.
3. El algoritmo es válido sin el supuesto (usado en el caso del costo esperado) de que la cadena de Markov asociada con toda matriz de transición es irreducible.

En el IOR Tutorial se incluye una rutina interactiva para aplicar este algoritmo.

Solución del ejemplo prototipo mediante este algoritmo de mejoramiento de la política. Se retoma el ejemplo prototipo en el punto que estaba antes de resumir este algoritmo.

Ya se seleccionó la política óptima de acuerdo con el criterio del costo promedio como la política inicial de prueba R_1. A continuación se resume esta política, su matriz de transición y sus costos.

Política R_1

Estado	Decisión
0	1
1	1
2	2
3	3

Matriz de transición

Estado	0	1	2	3
0	0	$\frac{7}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	1	0	0
3	1	0	0	0

Costos

Estado	C_{ik}
0	0
1	1 000
2	4 000
3	6 000

Además, ya se llevó a cabo el paso 1 (determinación del valor) de la iteración 1. Esta matriz de transición y estos costos condujeron a las ecuaciones empleadas para encontrar $V_0(R_1) = 14\,949$, $V_1(R_1) = 16\,262$, $V_2(R_1) = 18\,636$ y $V_3(R_1) = 19\,454$.

Para comenzar el paso 2 (mejoramiento de la política) sólo es necesario construir la expresión que debe minimizarse para los dos estados (1 y 2) con una decisión seleccionada.

Estado 1: $C_{1k} + 0.9[p_{10}(k)(14\,949) + p_{11}(k)(16\,262) + p_{12}(k)(18\,636) + p_{13}(k)(19\,454)]$

Estado 2: $C_{2k} + 0.9[p_{20}(k)(14\,949) + p_{21}(k)(16\,262) + p_{22}(k)(18\,636) + p_{23}(k)(19\,454)]$.

Para cada uno de estos estados y sus decisiones posibles, se muestran las C_{ik}, las $p_{ij}(k)$ y los valores que resultan de las expresiones correspondientes.

Decisión	Estado 1						
	C_{1k}	$p_{10}(k)$	$p_{11}(k)$	$p_{12}(k)$	$p_{13}(k)$	Valor de la expresión	
1	1 000	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	16 262	← Mínimo
3	6 000	1	0	0	0	19 454	

Decisión	Estado 2						
	C_{2k}	$p_{20}(k)$	$p_{21}(k)$	$p_{22}(k)$	$p_{23}(k)$	Valor de la expresión	
1	3 000	0	0	$\frac{1}{2}$	$\frac{1}{2}$	20 140	
2	4 000	0	1	0	0	18 636	← Mínimo
3	6 000	1	0	0	0	19 454	

Como la decisión 1 minimiza la expresión del estado 1 y la decisión 2 minimiza la expresión del estado 2, la próxima política de prueba (R_2) es la siguiente:

Política R_2

Estado	Decisión
0	1
1	1
2	2
3	3

Como esta política es idéntica a la política R_1, la prueba de optimalidad indica que esta política es óptima. Entonces, la política óptima según el criterio del costo esperado también es óptima de acuerdo con el criterio del costo descontado en este caso. (Esta situación ocurre con frecuencia, pero no siempre.)

Formulación de programación lineal

La formulación de programación lineal en el caso del costo descontado es similar a la del costo esperado que se presentó en la sección 19.3. Sin embargo, ya no se necesita la primera restricción que se dio en esa sección; las otras restricciones funcionales sí deben incluir el factor de descuento α. Otra diferencia es que el modelo ahora contiene constantes β_j, para $j = 0, 1, \ldots, M$. Estas constantes deben satisfacer las condiciones

$$\sum_{j=0}^{M} \beta_j = 1, \qquad \beta_j > 0 \qquad \text{para } j = 0, 1, \ldots, M,$$

excepto por esto, se puede elegir de manera arbitraria sin afectar la política óptima que se obtiene del modelo.

Con el modelo que resulta se debe elegir los valores de las variables de decisión *continuas* y_{ik} para

$$\text{Minimizar} \qquad Z = \sum_{i=0}^{M} \sum_{k=1}^{K} C_{ik} y_{ik},$$

sujeta a las restricciones

$$(1) \qquad \sum_{k=1}^{K} y_{jk} - \alpha \sum_{i=0}^{M} \sum_{k=1}^{K} y_{ik} p_{ij}(k) = \beta_j, \qquad \text{para } j = 0, 1, \ldots, M,$$

$$(2) \qquad y_{ik} \geq 0, \qquad \text{para } i = 0, 1, \ldots, M;\ k = 1, 2, \ldots, K.$$

Una vez que se usa el método símplex para obtener una solución óptima de este modelo, la política óptima correspondiente se define por

$$D_{ik} = P\{\text{decisión} = k \text{ y estado} = i\} = \frac{y_{ik}}{\sum_{k=1}^{K} y_{ik}}.$$

Ahora, las y_{ik} se pueden interpretar como un tiempo esperado *descontado* por estar en el estado i y tomar la decisión k, cuando la distribución de probabilidad del *estado inicial* (cuando comienzan las observaciones) es $P\{X_0 = j\} = \beta_j$ para $j = 0, 1, \ldots, M$. En otras palabras, si

$$z_{ik}^n = P\{\text{en el tiempo } n, \text{ estado} = i \text{ y decisión} = k\},$$

entonces

$$y_{ik} = z_{ik}^0 + \alpha z_{ik}^1 + \alpha^2 z_{ik}^2 + \alpha^3 z_{ik}^3 + \cdots.$$

Con la interpretación de las β_j como *probabilidades del estado inicial* (donde cada probabilidad es mayor que cero), Z se puede interpretar como el costo descontado total esperado correspondiente. En consecuencia, la elección de las β_j afecta el valor óptimo de Z (pero no la política óptima resultante).

Una vez más, es posible demostrar que la política óptima que se obtiene al resolver el modelo de programación lineal es determinística; es decir, $D_{ik} = 0$ o 1. Más aún, esta técnica es válida sin el supuesto (usado para el caso del costo esperado) de que la cadena de Markov asociada con cada matriz de transición es irreducible.

Solución del ejemplo prototipo mediante programación lineal. El modelo de programación lineal del ejemplo prototípico (con $\alpha = 0.9$) es

$$\text{Minimizar} \qquad Z = 1\,000 y_{11} + 6\,000 y_{13} + 3\,000 y_{21} + 4\,000 y_{22} + 6\,000 y_{23} + 6\,000 y_{33},$$

sujeta a

$$y_{01} - 0.9(y_{13} + y_{23} + y_{33}) = \frac{1}{4}$$

$$y_{11} + y_{13} - 0.9\left(\frac{7}{8}y_{01} + \frac{3}{4}y_{11} + y_{22}\right) = \frac{1}{4}$$

$$y_{21} + y_{22} + y_{23} - 0.9\left(\frac{1}{16}y_{01} + \frac{1}{8}y_{11} + \frac{1}{2}y_{21}\right) = \frac{1}{4}$$

$$y_{33} - 0.9\left(\frac{1}{16}y_{01} + \frac{1}{8}y_{11} + \frac{1}{2}y_{21}\right) = \frac{1}{4}$$

y

$$\text{toda } y_{ik} \geq 0,$$

donde β_0, β_1, β_2 y β_3 se seleccionan de manera arbitraria como $\frac{1}{4}$. El método símplex proporciona la solución óptima

$$y_{01} = 1.210, \quad (y_{11}, y_{13}) = (6.656, 0), \quad (y_{21}, y_{22}, y_{23}) = (0, 1.067, 0),$$
$$y_{33} = 1.067,$$

de manera que

$$D_{01} = 1, \quad (D_{11}, D_{13}) = (1, 0), \quad (D_{21}, D_{22}, D_{23}) = (0, 1, 0), \quad D_{33} = 1.$$

Esta política óptima es la misma que se obtuvo antes en esta sección mediante el algoritmo de mejoramiento de la política.

El valor de la función objetivo de la solución óptima es $Z = 17\,325$. Este valor está muy relacionado con los valores de las $V_i(R)$ de esta política óptima que se encontró mediante el algoritmo de mejoramiento de la política. Recuerde que cada $V_i(R)$ se interpreta como el costo descontado total esperado dado que el sistema se inicia en el estado i, y β_i se interpreta como la probabilidad de comenzar en el estado i. Como cada β_i se eligió igual a $\frac{1}{4}$,

$$17\,325 = \frac{1}{4}[V_0(R) + V_1(R) + V_2(R) + V_3(R)]$$
$$= \frac{1}{4}(14\,949 + 16\,262 + 18\,636 + 19\,454).$$

Procesos de decisión markovianos de periodo finito y el método de aproximaciones sucesivas

A continuación se dirigirá la atención hacia un enfoque llamado *método de aproximaciones sucesivas* para encontrar con *rapidez*, al menos, una *aproximación* a la política óptima.

Se ha supuesto que el proceso de decisión markoviano operará de manera indefinida y se ha buscado una política óptima para tal proceso. La idea básica del método de aproximaciones sucesivas es encontrar una política óptima para las decisiones que se toman en el primer periodo cuando sólo quedan n periodos de operación del proceso antes de terminar, comenzando con $n = 1$, después $n = 2$, después $n = 3$, etc. A medida que n crece, las políticas óptimas correspondientes convergen hacia la política óptima para el problema de periodo infinito de interés. Entonces, las políticas que se obtiene para $n = 1, 2, 3, \ldots$ proporcionan aproximaciones sucesivas que conducen a la política óptima deseada.

La razón por la que este enfoque es atractivo es que se cuenta con un método rápido para encontrar una política óptima cuando sólo quedan n periodos de operación, a saber, el de programación dinámica probabilística que se describió en la sección 10.4.

En particular, para $i = 0, 1, \ldots, M$, sea

V_i^n = costo descontado total esperado por seguir una política óptima, dado que el proceso comienza en el estado i y le quedan sólo n periodos de operación.[5]

[5] Como se quiere permitir que n crezca indefinidamente, se define n como el *número de periodos que quedan*, en lugar del *número de periodos desde el inicio* (como en el capítulo 10).

Según el *principio de optimalidad* de programación dinámica (vea la sección 10.2), las V_i^n se obtienen de la relación recursiva

$$V_i^n = \min_k \left\{C_{ik} + \alpha \sum_{j=0}^{M} p_{ij}(k)V_j^{n-1}\right\}, \qquad \text{para } i = 0, 1, \ldots, M.$$

El valor de k que minimiza proporciona la decisión óptima que se debe tomar en el primer periodo cuando el proceso se inicia en el estado i.

Para comenzar, con $n = 1$, todas las $V_i^0 = 0$, de manera que

$$V_i^1 = \min_k \{C_{ik}\}, \qquad \text{para } i = 0, 1, \ldots, M.$$

Aunque es posible que el método de aproximaciones sucesivas no conduzca a una política óptima en el caso del problema de periodos infinitos después de sólo unas cuantas iteraciones, tiene una ventaja especial sobre las técnicas de mejoramiento de la política y de programación lineal: nunca requiere que se resuelva un sistema de ecuaciones simultáneas, con lo que cada iteración se puede realizar en forma sencilla y rápida.

Aún más, si en realidad sólo quedan n periodos en el proceso de decisión markoviano, en definitiva n iteraciones de este método conducirán a una política óptima. (En el caso de un problema de n periodos, se permite establecer $\alpha = 1$, es decir, sin descuento, en cuyo caso el objetivo es minimizar el costo total esperado sobre los n periodos.)

En el IOR Tutorial se incluye una rutina interactiva como guía para utilizar este método de manera eficiente.

Solución del ejemplo prototipo por el método de aproximaciones sucesivas

Se usará de nuevo $\alpha = 0.9$. Consulte los valores de C_{ik} en la última columna de la tabla 19.1 al final de la sección 19.1. Vea también en las dos primeras columnas de la misma tabla que las únicas decisiones k factibles para cada estado i son $k = 1$, para $i = 0$, $k = 1$ o 3 para $i = 1$, $k = 1$, 2 o 3 para $i = 2$, y $k = 3$ para $i = 3$.

Para la primera iteración ($n = 1$), se muestra el valor que se obtuvo para cada V_i^1 junto con el valor de k que minimiza (dado entre paréntesis).

$$V_0^1 = \min_{k=1} \{C_{0k}\} = 0 \qquad (k = 1)$$

$$V_1^1 = \min_{k=1,3} \{C_{1k}\} = 1\,000 \qquad (k = 1)$$

$$V_2^1 = \min_{k=1,2,3} \{C_{2k}\} = 3\,000 \qquad (k = 1)$$

$$V_3^1 = \min_{k=3} \{C_{3k}\} = 6\,000 \qquad (k = 3)$$

En este contexto, la primera aproximación dice que se tome la decisión 1 (no hacer nada) cuando el sistema está en el estado 0, 1 o 2. Cuando el sistema se encuentra en el estado 3, se toma la decisión 3 (reemplazar la máquina).

La segunda iteración conduce a

$$V_0^2 = 0 + 0.9\left[\frac{7}{8}(1\,000) + \frac{1}{16}(3\,000) + \frac{1}{16}(6\,000)\right] = 1\,294 \qquad (k = 1)$$

$$V_1^2 = \min\left\{1\,000 + 0.9\left[\frac{3}{4}(1\,000) + \frac{1}{8}(3\,000) + \frac{1}{8}(6\,000)\right],\ 6\,000 + 0.9[1(0)]\right\} = 2\,688 \qquad (k = 1)$$

$$V_2^2 = \min\left\{3\,000 + 0.9\left[\frac{1}{2}(3\,000) + \frac{1}{2}(6\,000)\right],\ 4\,000 + 0.9[1(1\,000)],\ 6\,000 + 0.9[1(0)]\right\} = 4\,900 \qquad (k = 2)$$

$$V_3^2 = 6\,000 + 0.9[1(0)] = 6\,000 \qquad (k = 3).$$

donde el operador *min* se eliminó en la primera y cuarta expresiones debido a que había sólo una alternativa. La segunda aproximación indica que se deje la máquina como está cuando se encuentra en los estados 0 o 1, se haga una reparación general si está en el estado 2 y se reemplace cuando está en el estado 3. Observe que esta política es la óptima para el problema con número infinito de periodos, como se encontró en esta sección mediante el algoritmo de mejoramiento de la política y con programación lineal. Sin embargo, los valores de las V_i^2 (costo descontado total esperado cuando se comienza en el estado *i* para el problema de dos periodos) aún no son cercanos al de V_i (el costo correspondiente al problema del número infinito de periodos).

La tercera iteración da como resultado

$$V_0^3 = 0 + 0.9\left[\frac{7}{8}(2\,688) + \frac{1}{16}(4\,900) + \frac{1}{16}(6\,000)\right] = 2\,730 \qquad (k = 1)$$

$$V_1^3 = \min\left\{1\,000 + 0.9\left[\frac{3}{4}(2\,688) + \frac{1}{8}(4\,900) + \frac{1}{8}(6\,000)\right],\right.$$
$$\left. 6\,000 + 0.9[1(1\,294)]\right\} = 4\,041 \qquad (k = 1)$$

$$V_2^3 = \min\left\{3\,000 + 0.9\left[\frac{1}{2}(4\,900) + \frac{1}{2}(6\,000)\right],\right.$$
$$\left. 4\,000 + 0.9[1(2\,688)],\ 6\,000 + 0.9[1(1\,294)]\right\} = 6\,419 \qquad (k = 2)$$

$$V_3^3 = 6\,000 + 0.9[1(1\,294)] = 7\,165 \qquad (k = 3).$$

De nuevo se obtiene la política óptima para el problema de periodos infinitos y los costos se acercan a los de la política óptima para ese problema. Este procedimiento puede continuar y V_0^n, V_1^n, V_2^n y V_3^n y convergen a 14 949, 16 262, 18 636 y 19 454, respectivamente.

Debe observarse que de haber terminado el método de aproximaciones sucesivas después de la segunda iteración, se habría obtenido una política óptima para el problema de periodos infinitos, aunque no hay manera de saberlo si no se resuelve por otros métodos.

Como se indicó, en definitiva, el método de aproximaciones sucesivas obtiene una política óptima para un problema de *n* periodos después de *n* iteraciones. Para este ejemplo, la primera, segunda y tercera iteraciones han identificado la decisión óptima inmediata para cada estado si el número de periodos que quedan es uno, dos o tres, respectivamente.

19.6 CONCLUSIONES

Los procesos de decisión de Markov son una herramienta poderosa para optimizar el desempeño de los procesos estocásticos que se pueden modelar como cadena de Markov discreta. Las aplicaciones surgen en muchas áreas, como salud, mantenimiento de carreteras y puentes, administración de inventarios, mantenimiento de equipo, administración de flujos de efectivo, control de depósitos de agua, administración ambiental, control de sistemas de colas y operación de redes de comunicación. Las referencias seleccionadas 11 y 12 proporcionan una visión interesante de las primeras aplicaciones, la referencia 10 actualiza una aplicación que ganó un premio prestigiado, mientras que la referencia 4 describe otra aplicación laureada. Las referencias 3 y 8 incluyen información más reciente sobre aplicaciones.

Las dos medidas principales de desempeño que se usan son el *costo promedio esperado por unidad de tiempo* y el *costo descontado total esperado* (a largo plazo). La última medida requiere la determinación del valor adecuado de un factor de descuento, pero esta medida es útil cuando es importante tomar en cuenta el valor del dinero en el tiempo.

Los dos métodos más importantes para derivar políticas óptimas para los procesos de decisión markovianos son los *algoritmos de mejoramiento de una política* y *programación lineal*. Bajo el criterio de costo descontado, el *método de aproximaciones sucesivas* proporciona un camino rápido para aproximarse a una política óptima.

REFERENCIAS SELECCIONADAS

1. Altman, E.: *Constrained Markov Decision Processes*, Chapman and Hall/CRC, Boca Raton, FL, 1999.
2. Bertsekas, D. P.: *Dynamic Programming and Optimal Control*, vol. I, 2a. ed., Athena Scientific, Belmont MA, 2000.
3. Feinberg, E. A. y A. Shwartz: *Handbook of Markov Decision Processes: Method Applications*, Kluwer Academic Publishers, Boston, 2002.
4. Golabi, K. y R. Shepard: "Pontis: A System for Maintenance Optimization and Improvement of U.S. Bridge Networks", en *Interfaces*, **27**(1): 71-88, enero-febrero de 1997.
5. Howard R. A.: "Comments on the Origin and Application of Markov Decision Processes", en *Operations Research*, **50:** 100-102, enero-febrero de 2002.
6. Hu, J., M. C. Fu, V. R. Ramezani y S. I. Marcus: "An Evolutionary Random Policy Search Algorithm for Solving Markov Decision Processes", en *INFORMS Journal on Computing*, **19**(2): 161-174, primavera de 2007.
7. Puterman, M. L.: *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley, Nueva York, 1994.
8. Sennott, L. I.: *Stochastic Dynamic Programming and the Control of Queueing Systems*, Wiley, Nueva York, 1999.
9. Smith J. E. y K. F. McCardle: "Structural Properties of Stochastic Dynamic Programs", en *Operations Research*, **50:** 796-809, 2002.
10. Wang, K. C. P. y J. P. Zaniewski: "20/30 Hindsight: The New Pavement Optimization in the Arizona State Highway Network", en *Interfaces*, **26**(3): 77-89, mayo-junio de 1996.
11. White, D. J.: "Further Real Applications of Markov Decision Processes", en *Interfaces*, **18**(5): 55-61, septiembre-octubre de 1988.
12. White, D. J.: "Real Applications of Markov Decision Processes", en *Interfaces*, **15**(6): 73-83, noviembre-diciembre de 1985.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

Ejemplos resueltos

Ejemplos para el capítulo 19

Ejemplo de demostración en el OR Tutor:

Algoritmo de mejoramiento de la política: caso del costo promedio

Rutinas interactivas en IOR Tutorial:

Introducción del modelo de decisión de Markov
Algoritmo interactivo de mejoramiento de la política: costo promedio
Algoritmo interactivo de mejoramiento de la política: costo descontado
Método interactivo de aproximaciones sucesivas,

Rutinas automáticas en IOR Tutorial (Área de cadenas de Markov):

Introducción de la matriz de transición
Probabilidades de estado estable

Archivos para resolver formulaciones de programación lineal "Ch. 19—Markov Decision Proc"

Archivos de Excel
Archivo de LINGO/LINDO

Glosario para el capítulo 19

Vea la documentación del software en el apéndice 1.

PROBLEMAS

Los símbolos a la izquierda de algunos problemas (o de sus partes) significan lo siguiente:

D: El ejemplo de demostración indicado anteriormente puede ser útil.
I: Se sugiere el uso de las rutinas interactivas que se mencionaron con anterioridad (la impresión registra su trabajo).
A: Las rutinas automáticas que se indicaron con anterioridad pueden ser útiles.
C: Use la computadora con cualquier opción de software disponible (o la que le sugiera su instructor) para resolver la formulación de programación lineal.

Un asterisco en el número del problema indica que al final del libro se da al menos una respuesta parcial.

19.2-1. Lea el artículo de referencia que describe el estudio de investigación de operaciones que se resume en la Viñeta de aplicación que se presentó en la sección 19.2. Describa de manera breve la forma en que se aplicaron los procesos de decisión markovianos en este estudio. Después, elabore una lista de los beneficios financieros y no financieros que arrojó dicho estudio.

19.2-2.* En un periodo, un cliente potencial llega a una instalación de servicio con probabilidad de $\frac{1}{2}$. Si encuentra dos personas en ella (incluso la que es atendida en ese momento), el cliente potencial se retira de inmediato y nunca regresa; si hay una o menos, entra y se convierte en un cliente real. El administrador de la instalación dispone de dos tipos de configuraciones de servicio. Al principio de cada periodo debe decidir cuál de las dos usará. Si utiliza la configuración "lenta" con costo de 3 dólares y hay clientes presentes durante el periodo, el cliente que llega será atendido y se irá con probabilidad de $\frac{3}{5}$. Si utiliza la configuración "rápida" con costo de 9 dólares y hay clientes presentes durante el periodo, un cliente que llega será atendido y se irá con probabilidad de $\frac{4}{5}$. La probabilidad de que llegue más de un cliente o se sirva a más de uno en un periodo es cero. La ganancia es de 50 dólares por cliente atendido.

a) Formule este problema como un proceso de decisión de Markov. Identifique estados y decisiones. En cada combinación de estado y decisión, encuentre el *costo inmediato neto esperado* (reste la ganancia por servir al cliente) en que se incurre durante ese periodo.

b) Identifique todas las políticas (determinísticas estacionarias). Para cada una, elabore la matriz de transición y la expresión del costo neto esperado (a largo plazo) por periodo en términos de las probabilidades de estado estable desconocidas ($\pi_0, \pi_1, \ldots, \pi_M$).

A ***c***) Use el IOR Tutorial para encontrar las probabilidades de estado estable de cada política. Evalúe la expresión de *b*) y obtenga la política óptima por enumeración exhaustiva.

19.2-3.* Una estudiante está preocupada por su auto, pues no le gustan las abolladuras. Cuando maneja hasta la escuela puede estacionarlo en la calle en un espacio, en dos espacios o en el estacionamiento. En la calle, en un espacio, la probabilidad de que lo abollen es de $\frac{1}{10}$. En dos espacios es de $\frac{1}{50}$ y la probabilidad de una infracción de 15 dólares es de $\frac{3}{10}$. El estacionamiento le cuesta 5 dólares, pero su auto no sufrirá ningún daño. Si lo abollan y lo lleva a reparar, se queda sin auto 1 día y el costo asciende a 50 dólares por la reparación y el transporte en taxi. También puede manejar su auto abollado, pero piensa que la pérdida del valor y su orgullo equivale a un costo de 9 dólares por día de escuela. Desea determinar la política óptima para estacionarse y repararlo o no si lo abollan a fin de minimizar su costo promedio esperado (a largo plazo) por día de escuela.

a) Formule este problema como un proceso de decisión markoviano; identifique estados y decisiones y encuentre C_{ik}.

b) Identifique todas las políticas (determinísticas estacionarias). Para cada una, elabore la matriz de transición y escriba la expresión del costo promedio esperado (a la larga) por periodo en términos de las probabilidades de estado estable desconocidas ($\pi_0, \pi_1, \ldots, \pi_M$).

A ***c***) Use el IOR Tutorial para encontrar las probabilidades de estado estable de cada política. Evalúe la expresión que obtuvo en *b*) para encontrar la política óptima por enumeración exhaustiva.

19.2-4. Un hombre juega póker cada sábado en la noche en su casa con el mismo grupo de amigos. Si un sábado ofrece refrescos (con costo esperado de 14 dólares), el siguiente, el grupo tendrá una probabilidad de $\frac{7}{8}$ de jugar de buen humor y $\frac{1}{8}$ de hacerlo de mal humor. Si no lo hace, el siguiente sábado el grupo tendrá una probabilidad de jugar de buen humor de $\frac{1}{8}$ y de $\frac{7}{8}$ de hacerlo de mal humor, sin que importe el humor de este sábado. Es más, si la noche comienza de mal humor y él no ofrece refrescos, el grupo lo molestará y tendrá una pérdida esperada en el póker de 75 dólares. De otra manera, su promedio de ganancias o pérdidas es de cero. El hombre quiere encontrar la política óptima para determinar cuándo ofrecer refresco para minimizar su costo promedio esperado semanal (a largo plazo).

a) Formule este problema como un proceso de decisión markoviano; identifique estados y decisiones. Encuentre C_{ik}.

b) Identifique todas las políticas (determinísticas estacionarias). Para cada una, elabore la matriz de transición y obtenga la expresión del costo promedio esperado (a largo plazo) por periodo en términos de las probabilidades de estado estable desconocidas ($\pi_0, \pi_1, \ldots, \pi_M$).

A ***c***) Use el IOR Tutorial para encontrar las probabilidades de cada política. Evalúe la expresión que obtuvo en *b*) y obtenga la política óptima por enumeración exhaustiva.

19.2-5.* Cuando un jugador de tenis sirve, tiene dos oportunidades de servicio dentro del cuadro. Si falla las dos veces, pierde el punto. Si intenta un servicio as, sirve dentro del cuadro con probabilidad de $\frac{3}{8}$. Si saca un servicio suave, la probabilidad de hacerlo bien es $\frac{7}{8}$. Si sirve dentro del cuadro, gana el punto con probabilidad de $\frac{2}{3}$. Con el servicio suave dentro del cuadro, la probabilidad de ganar el punto es de $\frac{1}{3}$. Si el costo por cada punto perdido es $+1$ y por cada punto ganado es -1, el problema es determinar la estrategia óptima de servicio para minimizar (a largo plazo) el costo promedio esperado por punto. (*Sugerencia:* Sea el estado 0 un punto terminado, o cuando quedan dos servicios para el siguiente punto; sea el estado 1 el que denota un servicio faltante.)

a) Formule este problema como un proceso de decisión markoviano; identifique estados y decisiones. Encuentre C_{ik}.

b) Identifique todas las políticas (determinísticas estacionarias). Para cada una, elabore la matriz de transición y escriba la expre-

sión del costo promedio esperado (a largo plazo) del punto en términos de las probabilidades de estado estable desconocidas $(\pi_0, \pi_1, \ldots, \pi_M)$.

A ***c***) Use el IOR Tutorial para encontrar las probabilidades de cada política. Evalúe la expresión que obtuvo en *b*) para encontrar la política óptima por enumeración exhaustiva.

19.2-6. Cada año, la señora Fontanez puede invertir en dos fondos mutuos: el Go-Go Fund y el Go-Slow Fund. Al final de cada año, liquida su inversión, recoge sus ganancias y reinvierte. Las ganancias anuales de los fondos mutuos dependen de la reacción del mercado. En los últimos años el mercado ha oscilado alrededor de los 14 000 puntos, según las probabilidades de la siguiente matriz de transición:

	13 000	14 000	15 000
13 000	0.4	0.4	0.2
14 000	0.3	0.4	0.3
15 000	0.1	0.4	0.5

Cada año, si el mercado sube (o baja) 1 000 puntos, Go-Go Fund tiene ganancias (o pérdidas) de 25 000 dólares, mientras que las de Go-Slow Fund son de 10 000 dólares. Si un año el mercado sube (o baja) 2 000 puntos, las ganancias (o pérdidas) del Go-Go Fund serán de 60 000 dólares y las del Go-Slow Fund serán sólo de 25 000 dólares. Si el mercado no cambia, ninguno de los fondos tiene ganancias o pérdidas. La Sra. Fontanez quiere determinar su política óptima de inversión para minimizar su costo anual (pérdida menos ganancia) promedio esperado (a largo plazo).

a) Formule este problema como un proceso de decisión de Markov; identifique estados y decisiones. Calcule C_{ik}.

b) Identifique todas las políticas (determinísticas estacionarias). Para cada una, elabore la matriz de transición y escriba la expresión del costo promedio esperado (a la larga) por periodo en términos de las probabilidades de estado estable desconocidas $(\pi_0, \pi_1, \ldots, \pi_M)$.

A ***c***) Use el IOR Tutorial para determinar las probabilidades de cada política. Evalúe la expresión que obtuvo en *b*) para encontrar la política óptima por enumeración exhaustiva.

19.2-7. Buck y Bill Bogus son gemelos que trabajan en una gasolinera y tienen un negocio de falsificación en casa. Cada día deciden quién trabajará en la gasolinera, mientras que el otro se quedará en la imprenta del sótano de su casa. Se estima que cada día que la máquina trabaja bien producen 60 billetes de 20 dólares que se pueden usar. Sin embargo, la máquina no es confiable y se descompone seguido. Si no trabaja al iniciar el día, Buck la puede arreglar para el principio del día siguiente con probabilidad de 0.6. Si Bill trabaja en la máquina, la probabilidad disminuye a 0.5. Si Bill trabaja en ella cuando está operable tiene una probabilidad de 0.6 de que trabaje el día siguiente. Si Buck la trabaja, la probabilidad de que se descomponga es de 0.6. (Para simplificar, suponga que todas las descomposturas ocurren al final del día.) Los gemelos desean determinar la política óptima que les permita decidir cuándo quedarse cada uno en casa para maximizar (a largo plazo) la *ganancia* promedio esperada (cantidad de billetes falsos que se pueden usar) por día.

a) Formule el problema como un proceso de decisión de Markov; identifique estados y decisiones. Encuentre las C_{ik}.

b) Identifique todas las políticas (determinísticas estacionarias). Para cada una, elabore la matriz de transición y la expresión de la ganancia neta promedio esperada (a largo plazo) por periodo en términos de las probabilidades de estado estable desconocidas $(\pi_0, \pi_1, \ldots, \pi_M)$.

A ***c***) Use el IOR Tutorial para encontrar las probabilidades de cada política. Evalúe la expresión que obtuvo en *b*) para encontrar la política óptima por enumeración exhaustiva.

19.2-8. Considere un problema de inventarios de un solo producto con un número infinito de periodos, en el que al principio de cada uno de ellos se toma una decisión sobre el número de artículos que se producirán en él. El costo fijo es de 10 dólares y el costo unitario de producción de 5 dólares. El costo de mantener por cada artículo que no se vende durante el periodo es de 4 dólares (se puede guardar un *máximo* de 2 artículos). La demanda en cada periodo tiene distribución de probabilidad conocida: $\frac{1}{3}$ de probabilidad de que sean 0, 1 y 2 artículos, respectivamente. Si la demanda excede el inventario disponible en el periodo, las ventas se pierden y se incurre en un costo por faltantes (que incluye las ganancias perdidas) de 8 y 32 dólares por faltantes de 1 y 2 artículos, respectivamente.

a) Considere la política en la que se producen 2 artículos si al principio del periodo no hay existencias, y no se producen si hay cualquier número de artículos en el inventario. Determine el costo promedio esperado (a largo plazo) por periodo de esta política. Al encontrar la matriz de transición de la cadena de Markov, los estados representan los niveles de inventario al principio del periodo.

b) Identifique todas las políticas de inventario *factibles* (determinísticas estacionarias), es decir, las políticas que nunca conducen a superar la capacidad del almacén.

19.3-1. Reconsidere el problema 19.2-2.

a) Formule un modelo de programación lineal para encontrar la política óptima.

C ***b***) Utilice el método símplex para resolver el modelo. Use la solución óptima resultante para identificar la política óptima.

19.3-2.* Reconsidere el problema 19.2-3.

a) Formule un modelo de programación lineal para encontrar la política óptima.

C ***b***) Aplique el método símplex para resolver el modelo. Use la solución óptima resultante para identificar la política óptima.

19.3-3. Reconsidere el problema 19.2-4.

a) Formule un modelo de programación lineal para encontrar la política óptima.

C ***b***) Use el método símplex para resolver el modelo. Use la solución óptima resultante para identificar la política óptima.

19.3-4.* Reconsidere el problema 19.2-5.

a) Formule un modelo de programación lineal para encontrar la política óptima.

C ***b***) Aplique el método símplex para resolver el modelo. Utilice la solución óptima resultante para identificar la política óptima.

19.3-5. Reconsidere el problema 19.2-6.

a) Formule un modelo de programación lineal para encontrar la política óptima.

C ***b***) Utilice el método símplex para resolver el modelo. Use la solución óptima resultante para identificar la política óptima.

19.3-6. Reconsidere el problema 19.2-7.

a) Formule un modelo de programación lineal para encontrar la política óptima.

C ***b*)** Utilice el método símplex para resolver el modelo. Use la solución óptima resultante para identificar la política óptima.

19.3-7. Reconsidere el problema 19.2-8.

***a*)** Formule un modelo de programación lineal para encontrar la política óptima.

C ***b*)** Aplique el método símplex para resolver el modelo. Utilice la solución óptima resultante para identificar la política óptima.

D,I **19.4-1.** Con el algoritmo de mejoramiento de políticas encuentre la política óptima para el problema 19.2-2.

D,I **19.4-2.*** Con el algoritmo de mejoramiento de políticas encuentre la política óptima para el problema 19.2-3.

D,I **19.4-3.** Con el algoritmo de mejoramiento de políticas encuentre la política óptima para el problema 19.2-4.

D,I **19.4-4.*** Con el algoritmo de mejoramiento de políticas encuentre la política óptima para el problema 19.2-5.

D,I **19.4-5.** Con el algoritmo de mejoramiento de políticas encuentre la política óptima para el problema 19.2-6.

D,I **19.4-6.** Con el algoritmo de mejoramiento de políticas encuentre la política óptima para el problema 19.2-7.

D,I **19.4-7.** Con el algoritmo de mejoramiento de políticas encuentre la política óptima para el problema 19.2-8.

D,I **19.4-8.** Considere el problema 16.5-5 del banco de sangre que se presentó en páginas anteriores. Suponga que el número de litros de sangre que recibe (en entrega normal) se puede especificar en el momento de la entrega (en lugar de la política de recibir 1 litro cada vez). Por lo tanto, el número de litros que recibe puede ser 0, 1, 2 o 3 (nunca se usan más de 3). El costo del reparto normal es de 50 dólares por litro, y el costo de la entrega urgente es de 100 dólares por litro. Inicie con la política de tomar un litro en cada entrega regular si el número de litros disponible justo antes de la entrega es de 0, 1 o 2 litros (es decir, nunca hay más de tres litros a la mano), y lleve a cabo dos iteraciones del algoritmo de la mejora de la política. (Debido a que muy pocos litros se tienen a la mano y los litros más antiguos siempre se usan primero, usted puede pasar por alto la remota posibilidad de que ninguno de los litros llegará a mantenerse en la repisa por 21 días y que sea necesario eliminarlos.)

I **19.5-1.*** Joe quiere vender su automóvil. Cada mes le hacen una oferta y debe decidir de inmediato si aceptarla o no. Una vez rechazada, pierde la oportunidad. Las ofertas posibles son de 600, 800 y 1 000 dólares, con probabilidades respectivas de $\frac{5}{8}$, $\frac{1}{4}$, y $\frac{1}{8}$ (las ofertas sucesivas son independientes entre sí). Existe un costo de mantenimiento de 60 dólares al mes. Joe está ansioso por vender el auto y ha elegido un factor de descuento $\alpha = 0.95$.

Utilice el algoritmo de mejoramiento de la política para encontrar la que minimiza el costo descontado total esperado. (*Sugerencia:* Se puede optar por dos acciones, esto es, aceptar o rechazar la oferta. Defina el espacio de estados como la oferta en el mes t. Incluya también el estado ∞, donde el proceso llega a este estado siempre que se acepte la oferta y permanece ahí con un costo mensual de 0.)

19.5-2.* Reconsidere el problema 19.5-1.

***a*)** Formule un modelo de programación lineal para encontrar la política óptima.

C ***b*)** Aplique el método símplex para resolver el modelo. Utilice la solución óptima resultante para identificar la política óptima.

I **19.5-3.*** En el caso del problema 19.5-1, utilice tres iteraciones del método de aproximaciones sucesivas para obtener una solución cercana a la óptima.

I **19.5-4.** El precio de cierta acción fluctúa entre 10, 20 y 30 dólares de un mes a otro. Los analistas de mercado han determinado que si el precio es de 10 dólares durante cualquier mes, el siguiente mes será de 10 o 20 dólares, con probabilidades respectivas de $\frac{4}{5}$ y $\frac{1}{5}$; si la acción está en 20 dólares, el siguiente mes tiene probabilidades de $\frac{1}{4}$, $\frac{1}{4}$ y $\frac{1}{2}$, de que el precio sea de 10, 20 o 30 dólares. Si la acción vale 30 dólares, el precio el siguiente mes será de 20 o 30 dólares con probabilidades respectivas de $\frac{3}{4}$ y $\frac{1}{4}$, Dado un factor de descuento de 0.9, use el algoritmo de mejoramiento de la política para determinar cuándo conviene vender y cuándo es mejor quedarse con las acciones, de manera que se maximice la ganancia descontada total esperada. (*Sugerencia:* Incluya un estado al que se llega con probabilidad 1 cuando se vende la acción y con probabilidad 0 cuando se conserva.)

19.5-5. Reconsidere el problema 19.5-4.

***a*)** Formule un modelo de programación lineal para encontrar una política óptima.

C ***b*)** Aplique el método símplex para resolver el modelo. Utilice la solución óptima para identificar la política óptima.

I **19.5-6.** En el caso del problema 19.5-4, realice tres iteraciones del método de aproximaciones sucesivas para obtener una solución cercana a la política óptima.

19.5-7. Una compañía química produce dos sustancias denotadas por C1 y C2, y sólo se puede producir una de ellas a la vez. Cada mes se toma la decisión de cuál de las dos producir. Como se puede predecir la demanda de cada sustancia, es posible saber que si se produce la 2 este mes, existe 60% de probabilidades de que también se produzca el mes próximo. De igual manera, si se produce la 1 este mes, sólo se tiene 30% de probabilidades de que se produzca de nuevo el mes próximo.

La compañía química tiene dos procesos para combatir la emisión de contaminantes; el proceso A es eficiente *sólo* para combatir la contaminación que proviene de la producción de la sustancia 2, mientras que el proceso B es eficiente *sólo* para combatir la que se debe a la producción de la sustancia 1. Únicamente se puede usar un proceso a la vez. La cantidad de contaminantes que surgen de la producción de cada sustancia en ambos procesos es

	C1	C2
A	15	2
B	3	8

Desafortunadamente, el tiempo de preparación de los procesos de control de contaminación causa un retraso, por lo que la decisión de cuál de ellos se utilizará debe tomarse un mes antes. La administración quiere determinar una política para decidir cuándo debe usar cada proceso de control de contaminación para minimizar la cantidad descontada total de contaminación futura con un factor de descuento $\alpha = 0.5$.

***a*)** Formule este problema como un proceso de decisión markoviano; identifique estados, decisiones y las C_{ik}. Identifique todas las políticas (determinísticas estacionarias).

I ***b*)** Utilice el algoritmo de mejoramiento de una política para encontrar la política óptima.

19.5-8. Reconsidere el problema 19.5-7.

***a*)** Formule un modelo de programación lineal para encontrar la política óptima.

C ***b*)** Utilice el método símplex para resolver el modelo. Emplee la solución óptima resultante para identificar la política óptima.

I **19.5-9.** En el problema 19.5-7 realice dos iteraciones del método de aproximaciones sucesivas para encontrar una política cercana a la óptima.

I **19.5-10.** Reconsidere el problema 19.5-7. Suponga que la compañía producirá cualquiera de las sustancias sólo durante 4 meses más, de manera que la decisión respecto del proceso de control de contaminación dentro de 1 mes sólo tiene que tomarse tres veces más. Encuentre la política óptima para este problema de tres periodos.

I **19.5-11.*** Reconsidere el ejemplo prototipo de la sección 19.1. Suponga que el proceso de producción que utiliza la máquina bajo consideración se usará sólo 4 semanas más. Encuentre la política óptima para este problema de cuatro periodos mediante el criterio del costo descontado con un factor de descuento $\alpha = 0.9$.

CAPÍTULO 20

Simulación

Este capítulo final se dedica a la última técnica importante de investigación de operaciones. La *simulación* se clasifica en un escalón muy alto entre las técnicas que más se usan. Aún más, debido a que es una herramienta tan flexible, poderosa e intuitiva, sus aplicaciones crecen con rapidez de manera continua.

Esta técnica involucra el uso de una computadora para *imitar* (simular) la operación de un proceso o sistema completo. Por ejemplo, a menudo se usa simulación para realizar un análisis de riesgo de procesos financieros mediante la imitación repetida de la evolución de las transacciones necesarias para generar un perfil de los resultados posibles. También se utiliza ampliamente en el análisis de sistemas estocásticos que continuarán en operación indefinidamente. En el caso de este tipo de sistemas, la computadora genera y registra las ocurrencias de los eventos que impulsan el sistema como si en realidad estuviera en operación física. Debido a su velocidad, la computadora puede simular incluso años de operación en cuestión de segundos. El registro del desempeño de la operación simulada del sistema de varias alternativas de diseño o procedimientos de operación permite evaluar y comparar estas alternativas antes de elegir una.

En la primera sección se describe e ilustra la esencia de la simulación. En la siguiente se presenta una variedad de aplicaciones comunes de esta técnica. En las secciones 20.3 y 20.4 se estudian dos herramientas clave: la generación de números aleatorios y la generación de observaciones aleatorias a partir de distribuciones de probabilidad. La sección 20.5 describe el procedimiento global para simular. La siguiente sección explica cómo es posible realizar con eficiencia algunas simulaciones en hojas de cálculo. Un complemento de este capítulo, que se encuentra en el sitio en internet de este libro, introduce algunas técnicas especiales para mejorar la precisión de las estimaciones de las medidas de desempeño del sistema que se simula. Un segundo complemento presenta un método estadístico innovador para analizar los resultados de una simulación. Un tercero extiende el enfoque basado en hojas de cálculo en la búsqueda de una solución óptima de los modelos de simulación.

20.1 ESENCIA DE LA SIMULACIÓN

Desde hace mucho tiempo, la técnica de *simulación* ha sido una herramienta importante para el diseñador. Por ejemplo, la simulación del vuelo de un avión en un túnel de viento es una práctica normal con los nuevos diseños. En teoría, se podrían usar las leyes de la física para obtener la misma información sobre los cambios en el desempeño del avión si cambian los parámetros, pero en sentido práctico, el análisis sería muy complejo. Otra alternativa sería construir aviones reales para cada diseño y probarlos en vuelos reales para elegir el diseño final, pero este recurso sería demasiado costoso (al igual que peligroso). Por lo tanto, después de realizar un análisis teórico preliminar para desarrollar un diseño *básico*, la herramienta viable para experimentar con los diseños *específicos* es la simulación del vuelo en un túnel de viento. Esta actividad implica *imitar* el desempeño de un avión real en un medio controlado con el fin de *estimar* cuál sería su desempeño real. Después de

desarrollar un diseño detallado de esta manera, es posible construir un prototipo y probarlo en un vuelo real para ajustar los últimos detalles del diseño final.

Papel de la simulación en los estudios de investigación de operaciones

En esencia, la *simulación* tiene el mismo papel en muchos estudios de IO. No obstante, en lugar de diseñar un avión, el equipo de IO se dedica a desarrollar un diseño o procedimiento de operación para algún *sistema estocástico* (que opera en forma *probabilística* a través del tiempo). Algunos de estos sistemas estocásticos recuerdan los ejemplos de cadenas de Markov y líneas de espera que se describieron en los capítulos 16 y 17, mientras que otros son más elaborados. En lugar de usar un túnel de viento, el desempeño del sistema real se *imita* mediante distribuciones de probabilidad para *generar aleatoriamente* los distintos eventos que ocurren en el sistema. Por todo esto, un modelo de simulación *sintetiza* el sistema con su construcción de cada componente y de cada evento. Después, el modelo *corre* el sistema simulado para obtener *observaciones estadísticas* del desempeño del sistema como resultado de los diferentes eventos generados de manera aleatoria. Debido a que las *corridas de simulación*, por lo general, requieren la generación y el procesado de una gran cantidad de datos, es inevitable que estos experimentos estadísticos simulados se lleven a cabo en una computadora.

Cuando es necesario usar simulación como parte de un estudio de IO, es común que vaya precedida y seguida de los mismos pasos que se describieron antes para diseñar un avión. En particular, primero se hace un análisis teórico preliminar (quizá con modelos matemáticos aproximados) para desarrollar un diseño básico del sistema (que incluye sus procedimientos de operación). Después se usa simulación para experimentar con los diseños específicos con el fin de estimar el desempeño real. Una vez desarrollado y elegido el diseño detallado, se prueba el sistema real para ajustar los últimos detalles del diseño final.

Para preparar la simulación de un sistema complejo, es necesario contar con un **modelo de simulación** detallado para formular y describir la operación del sistema y cómo debe simularse, el cual consta de varios bloques de construcción básicos:

1. Definir el *estado del sistema* (como el número de clientes en un sistema de colas).
2. Identificar los *estados posibles* del sistema que pueden ocurrir.
3. Identificar los *eventos posibles* (como las llegadas y terminaciones de servicio en un sistema de colas) que cambian el estado del sistema.
4. Contar con un *reloj de simulación*, localizado en alguna dirección del programa de simulación, que registrará el paso del tiempo (simulado).
5. Un método para *generar los eventos de manera aleatoria* de los distintos tipos.
6. Una fórmula para identificar las *transiciones de los estados* que generan los diferentes tipos de eventos.

Se han logrado grandes progresos en el desarrollo de paquetes de software especiales (que se describirán en la sección 20.5) para integrar con eficiencia el modelo de simulación a un programa de computadora y realizar las simulaciones. De cualquier forma, cuando se trata de sistemas más o menos complejos, la simulación tiende a ser un procedimiento costoso. Después de formular un modelo detallado, con frecuencia se requiere mucho tiempo para desarrollar y depurar los programas de computadora para las corridas de simulación. Además, es posible que se requieran muchas corridas para obtener buenas *estimaciones* del desempeño de los diseños alternativos del sistema. Por último, todos los datos deben analizarse con cuidado antes de obtener conclusiones finales. Este proceso completo suele tomar mucho tiempo y esfuerzo. Por lo tanto, la simulación no debe ser utilizada cuando exista un procedimiento menos costoso que pueda generar la misma (o mejor) información.

La simulación casi siempre se usa cuando el sistema estocástico en cuestión es demasiado complejo para que su análisis con los modelos analíticos (como modelos de colas) descritos en los capítulos anteriores, sea satisfactorio. Lo más importante del enfoque analítico es que abstrae la esencia del problema, revela su estructura fundamental y proporciona una visión de las relaciones causa-efecto dentro del sistema. Por ello, si es posible construir un modelo analítico que sea, a la vez, una idealización razonable del problema y una solución satisfactoria, este enfoque es superior a la simulación. Sin embargo, muchos problemas son tan complejos que no se pueden resolver por

la vía analítica. Entonces, aunque la simulación tiende a ser un proceso relativamente caro, con frecuencia es el único enfoque práctico para resolver un problema.

Simulación de eventos discretos *versus* continuos

Las dos grandes categorías de simulación son la de eventos discretos y eventos continuos.

Cuando se recurre a una **simulación de eventos discretos**, los cambios en el estado del sistema ocurren de manera instantánea en puntos aleatorios del tiempo como resultado de la ocurrencia de *eventos discretos*. Por ejemplo, en un sistema de colas donde el estado del sistema es el número de clientes en él, los eventos discretos que cambian este estado son la llegada de un cliente o su salida cuando termina su servicio. En la práctica, la mayoría de las aplicaciones de simulación son simulaciones de eventos discretos.

En una **simulación continua** los cambios en el estado del sistema ocurren *continuamente* en el tiempo. Por ejemplo, si el sistema de interés es un avión en vuelo y su estado se define como la posición actual, el estado cambia de manera continua en el tiempo. Algunas aplicaciones de simulaciones continuas ocurren en los estudios de diseño de sistemas de ingeniería de este tipo. Las simulaciones continuas suelen requerir ecuaciones diferenciales para describir la tasa de cambio de las variables de estado, por lo que el análisis tiende a ser complejo.

En ocasiones es posible aproximar los cambios continuos en el estado del sistema mediante cambios discretos, para usar una simulación de eventos discretos que aproxime el comportamiento de un sistema continuo, circunstancia que tiende a simplificar mucho el análisis.

En adelante, este capítulo se concentrará en las simulaciones de eventos discretos. Se supondrá esta característica en todas las referencias subsecuentes de simulación.

A continuación se verán dos ejemplos para ilustrar las ideas básicas de simulación. Estos ejemplos se han mantenido mucho más sencillos que las aplicaciones usuales de esta técnica, con el fin de que sea más fácil resaltar las ideas primordiales. El primer sistema es tan sencillo que, en realidad, ni siquiera tiene que realizarse en una computadora. El segundo incorpora más de las características normales de una simulación, aunque también es tan sencillo como para obtener una solución analítica.

EJEMPLO 1 Juego de lanzamiento de monedas

Usted es el afortunado ganador de un concurso. El premio es un viaje todo pagado a uno de los hoteles importantes de Las Vegas, que incluye algunas fichas para apostar en el casino del hotel.

Al entrar al casino, se da cuenta de que además de los juegos tradicionales (blackjack, ruleta, etc.) ofrecen un nuevo juego con las siguientes reglas.

Reglas del juego

1. En cada jugada se lanza una moneda no alterada en repetidas ocasiones hasta que la *diferencia* entre el número de caras y cruces que aparecen sea tres.
2. Si decide participar, debe pagar un dólar cada vez que se lanza la moneda. No puede abandonar el juego hasta que éste acaba.
3. Se reciben 8 dólares al final de cada juego.

En consecuencia, se gana dinero si el número de lanzamientos es menor que ocho, pero se pierde si se tiene que lanzar la moneda más de ocho veces. Éstos son algunos ejemplos (donde H representa cara y T cruz).

HHH	3 lanzamientos.	Se gana $5
THTTT	5 lanzamientos.	Se gana $3
THHTHTHTTTT	11 lanzamientos.	Se pierde $3

¿Cómo se podría decidir si conviene o no participar en este juego?

Muchas personas basarían esta decisión en la *simulación*, aunque tal vez no la llamaran así. En este caso, la simulación no es más complicada que jugar uno mismo el juego muchas veces hasta que sea claro si vale la pena jugarlo por dinero. Podría bastar con lanzar una moneda durante media

hora y registrar las pérdidas y ganancias que resultan. En realidad, ésta es una simulación porque *imita* el juego real *sin* que en realidad se gane o se pierda dinero.

Ahora se verá cómo se puede usar una computadora para realizar este mismo *experimento simulado*. Aunque una computadora no puede lanzar monedas, puede *simular* que lo hace. Logra esto mediante la generación de una secuencia de *observaciones aleatorias* de una distribución uniforme entre 0 y 1, donde se hace referencia a estas observaciones aleatorias como *números aleatorios uniformes* en el intervalo [0, 1]. Una manera sencilla de generar números aleatorios uniformes es usar la función **ALEATORIO**() de Excel. Por ejemplo, la esquina inferior izquierda de la figura 20.1 indica que se introdujo =ALEATORIO() en la celda C13 y después se copió en el rango C14:C62 con el comando Copiar (los paréntesis deben incluirse en esta función, aunque no se escriba algo entre ellos). Esto ocasiona que Excel genere los números aleatorios que se muestran en las celdas C13:C62 de la hoja de cálculo (las filas 27 a 56 están escondidas para ahorrar espacio en la figura).

Las probabilidades del resultado de lanzar la moneda son

$$P(\text{caras}) = \frac{1}{2}, \qquad P(\text{cruces}) = \frac{1}{2}.$$

Por lo tanto, para simular el lanzamiento de una moneda en la computadora *cualquier mitad* de los números aleatorios posibles corresponden a *cara* y la *otra mitad* a *cruz*. Para ser específicos, se usará la siguiente correspondencia:

0.0000 a 0.4999	corresponde a	*cara*.
0.5000 a 0.9999	corresponde a	*cruz*.

Al usar la fórmula

$$= \text{SI(Aleatorio número} < 0.5, \text{“Cara”, “Cruz”)},$$

en cada celda de la columna D de la figura 20.1, Excel inserta Cara si el número aleatorio es menor que 0.5 y Cruz de otra manera. En consecuencia, los primeros 11 números aleatorios generados en la columna C llevan a la siguiente secuencia de caras (H) y cruces (T):

HTTTHHHTHHH,

punto en el que se detiene el juego porque el número de cruces (7) excede por 3 al número de caras (4). Las celdas D7 y D8 registran el número total de lanzamientos (11) y las ganancias que se obtienen ($8 – $11 = –$3).

Las ecuaciones de la parte inferior de la figura 20.1 muestran las fórmulas que se introdujeron en las celdas de la parte superior y que después se copiaron en otra hoja. Si se usan estas ecuaciones, la hoja de cálculo registra la simulación de una jugada completa del juego. Para asegurar que el juego termine, se simularon 50 lanzamientos. Las columnas E y F registran el número acumulado de caras y cruces cada vez que se lanza. Las ecuaciones de las celdas de la columna G dejan la celda en blanco hasta que la diferencia entre el número de caras y cruces llega a 3, en cuyo punto se inserta un STOP en la celda. De ahí en adelante aparece NA (no se aplica). Mediante el empleo de las ecuaciones que se muestran en la parte superior derecha de la figura 20.1 se registra en las celdas D7 y D8 el resultado de la jugada simulada del juego.

Estas simulaciones del juego de monedas se pueden repetir cuantas veces se desee con esta hoja de cálculo. Cada vez, Excel genera una nueva secuencia de números aleatorios y, por ende, de caras y cruces. (Excel repetirá una secuencia de números aleatorios sólo si se elige el intervalo de números que se desea repetir, copiándolos como valores con el “pegado especial” del menú de “edición”.)

Es común que las simulaciones se repitan muchas veces para obtener una estimación más confiable con un resultado promedio. En consecuencia, esta misma hoja se usa para generar los datos de la tabla de la figura 20.2 para 14 jugadas. Como se indica en la parte superior derecha de la figura 20.2, esto se hace mediante la introducción de ecuaciones en el primer renglón de la tabla que se refiere a las celdas de salida de la figura 20.1, es decir, se escribe = NumeroLanzamientos en la celda K6 e = Ganados en la celda L6, dejando la celda J6 en blanco. El siguiente paso es seleccionar el contenido completo de la tabla (celdas J6:L20) y elegir Data Table del menú

	A	B	C	D	E	F	G
1	**Juego de lanzamiento de moneda**						
2							
3		Diferencia requerida		3			
4		Efectivo al final del juego		$8			
5							
6				**Resumen del juego**			
7		Número de lanzamientos		11			
8		Ganados		-$3			
9							
10							
11		Lanza-	Número		Total	Total	
12		miento	aleatorio	Resultado	de caras	de cruces	¿Parar?
13		1	0.6961	Cara	1	0	
14		2	0.2086	Cruz	1	1	
15		3	0.1457	Cruz	1	2	
16		4	0.3098	Cruz	1	3	
17		5	0.6996	Cara	2	3	
18		6	0.9617	Cara	3	3	
19		7	0.6117	Cara	4	3	
20		8	0.3948	Cruz	4	4	
21		9	0.7769	Cara	5	4	
22		10	0.5750	Cara	6	4	
23		11	0.6271	Cara	7	4	Parar
24		12	0.2017	Cruz	7	5	NA
25		13	0.7660	Cara	8	5	NA
26		14	0.9918	Cara	9	5	NA
57		45	0.2461	Cruz	23	22	NA
58		46	0.7011	Cara	24	22	NA
59		47	0.3533	Cruz	24	23	NA
60		48	0.7136	Cara	25	23	NA
61		49	0.7876	Cara	26	23	NA
62		50	0.3580	Cruz	26	24	NA

Nombre de rango	Celdas
EfectivoAlFinalDelJuego	D4
Lanzamiento	B13:B62
Número de Lanzamientos	D7
Número Aleatorio	C13:C62
Diferencia Requerida	D3
Resultado	D13:D62
¿Parar?	G13:G62
Total de Caras	E13:E62
Total de Cruces	F13:F62
Ganados	D8

	C	D
6		**Resumen del juego**
7	Núm. de lanz.	=CONTARBLANCOS (¿parar) + 1
8	Ganados	=Efectivo al final del juego

	C	D	E	F
11	Número		Total	Total
12	Aleatorio		de caras	de cruces
13	=ALEAT.()	=SI(NúmeroAleatorio<0.5,1,0)	=SI(Resultado=”Cara”,1,0)	=Lan.-Total de Caras
14	=ALEAT.()	=SI(NúmeroAleatorio<0.5,”Cruz”,”Cara”)	=E13+=SI(Resultado=”Cara”,1,0)	=Lan.-Total de Caras
15	=ALEAT.()	=SI(NúmeroAleatorio<0.5,”Cruz”,”Cara”)	=E14+=SI(Resultado=”Cara”,1,0)	=Lan.-Total de Caras
16	:	:	:	:
17	:	:	:	:

	G
12	¿Parar?
13	
14	
15	=SI(ABS(TotalDeCaras-TotalDeCruces)>=DiferenciaRequerida,”Parar”,””)
16	=SI(G15=””,SI(ABS(TotalDeCaras-TotalDeCruces)>=DiferenciaRequerida,”Parar”,””),”NA”)
17	=SI(G16=””,SI(ABS(TotalDeCaras-TotalDeCruces)>=DiferenciaRequerida,”Parar”,””),”NA”)
18	:
19	:

■ **FIGURA 20.1**
Modelo en hoja de cálculo de una simulación del juego de lanzamiento de monedas (ejemplo 1).

	I	J	K	L	M
1	**Tabla de datos del juego de lanza-**				
2	**miento de monedas (14 réplicas)**				
3					
4			Número de		
5		Jugada	lanzams.	Ganados	
6			3	$5	Seleccione toda la tabla (J6:L20) antes de elegir "Tabla" del menú de datos.
7		1	9	-$1	
8		2	5	$3	
9		3	7	$1	
10		4	11	-$3	
11		5	5	$3	
12		6	3	$5	
13		7	3	$5	
14		8	11	-$3	
15		9	7	$1	
16		10	15	-$7	
17		11	3	$5	
18		12	7	$1	
19		13	9	-$1	
20		14	5	3	
21					
22		Promedio	7.14	$0.86	

Nombre de rango	Celda
Número de Lan.	D7
Ganados	D8

	K	L
4	Número	
5	de lanzamientos	Ganados
6	Numero de Lan.	=Ganados



	J	K	L
22	Promedio	=PROMEDIO(K7:K20)	=PROMEDIO(L7:L20)

■ **FIGURA 20.2**
Tabla de datos que registra los resultados de 14 réplicas de la simulación con hoja de cálculo de la figura 20.1.

What-IfAnalysis del Data tab (en el caso de Excel 2007) o Table del menú Data (en el caso de versiones anteriores de Excel). Por último se selecciona *cualquier* celda en blanco (por ejemplo, E4) de la columna de celdas de entrada y se hace clic en aceptar. Excel ingresa los números en la primera columna de la tabla (J7:J20) y usa las celdas C13:G62 de toda la hoja de cálculo original (figura 20.1) para calcular de nuevo las celdas de salida de las columnas K y L para cada renglón en el que se coloca *cualquier* número en la columna J. Después se introducen las ecuaciones =PROMEDIO(K7:K20) o (L7:L20) en las celdas K22 y L22 para obtener los promedios.

Aunque esta corrida de simulación en particular requirió de dos hojas de cálculo, una para realizar las réplicas de la simulación y la otra para registrar los resultados de las réplicas en la tabla, debe señalarse que las réplicas de algunas otras simulaciones se pueden realizar en una sola hoja. Éste es el caso siempre que cada réplica se pueda realizar y registrar en un renglón de la hoja de cálculo. Por ejemplo, si sólo se necesita un número aleatorio uniforme para realizar la réplica, entonces toda la corrida completa de simulación se puede realizar y registrar en una hoja de cálculo similar a la de la figura 20.1.

De regreso a la figura 20.2, la celda K22 indica que esta muestra de 14 jugadas da un promedio muestral de 7.14 lanzamientos. El promedio muestral proporciona una *estimación de la media* verdadera de la distribución de probabilidad que sigue el número de lanzamientos que se requiere para una jugada del juego. Por ello, el promedio muestral de 7.14 parecería indicar que, en promedio, debe ganar alrededor de 0.86 de dólar (celda L22) cada vez que juegue. Por lo tanto, si no se tiene una aversión alta al riesgo, parece que debe elegirse jugar, de preferencia un número grande de veces.

Pero ¡*cuidado*! Un error común en el uso de simulación es que las conclusiones se basen en muestras demasiado pequeñas o porque se hizo un análisis estadístico inadecuado o se carece de él por completo. En este caso, la *desviación estándar de la muestra* es 3.67, de manera que la *desviación estándar* estimada del *promedio de la muestra* es $3.67/\sqrt{14} \approx 0.98$. Por ello, aun cuando se suponga que la distribución de probabilidad del número de lanzamientos que se requieren en una jugada sigue una *distribución normal* (que es un supuesto con poco fundamento, dado que la distribución real es *sesgada*), cualquier *intervalo de confianza* razonable de la *media* verdadera de

esta distribución va más allá del 8. Así, para obtener una conclusión válida a un nivel razonable de significancia estadística, se requiere un tamaño de muestra mucho más grande. Desafortunadamente, como la desviación estándar de un promedio muestral es inversamente proporcional a la *raíz cuadrada* del tamaño de la muestra, se necesita un incremento considerable del tamaño de la muestra para obtener un pequeño aumento en la exactitud de la estimación de la media verdadera. En este caso, parece que sería adecuado realizar 100 jugadas simuladas (réplicas) del juego, según qué tan cercano sea el promedio a 8, pero mil réplicas sería mucho más seguro.

Ocurre que la *media* verdadera del número de lanzamientos que se requiere en una jugada es 9 (esta media se puede encontrar por vía analítica, pero no es sencillo). En consecuencia, a la larga, en realidad el promedio es perder 1 dólar cada juego. Parte de la razón para que el experimento simulado no llegue a esta conclusión es que existe una oportunidad muy pequeña de una pérdida grande en cualquier jugada, pero nunca se puede ganar más de 5 dólares cada vez. Sin embargo, 14 jugadas simuladas no fueron suficientes para obtener observaciones que se encontraran lejos, en la cola de la distribución de la cantidad ganada o perdida en una jugada. Sólo dos jugadas simuladas dieron una pérdida de más de 3 dólares, y fue sólo de 7.

En la figura 20.3 se presentan los resultados de correr la simulación de mil jugadas (donde las filas 17 a 1 000 no se muestran). La celda K1008 contiene el número promedio de lanzamientos como 8.97, muy cercano a la media verdadera de 9. Con este número de réplicas, la ganancia promedio de –$0.97 en la celda L1008 proporciona una base confiable para concluir que en este juego no se gana dinero a la larga (puede estar seguro que el casino ya usó simulación para verificar esta afirmación de antemano).

Aunque en realidad no era necesario construir un *modelo de simulación* completo para realizar esta sencilla simulación, se hace con fines ilustrativos. El *sistema estocástico* que se simula consiste en lanzamientos sucesivos de una moneda en cada jugada. El *reloj de la simulación* registra el número de lanzamientos (simulados) t que se ha realizado hasta ahora. La información sobre el sistema que define el estado actual, es decir, el *estado del sistema* es

$N(t)$ = número de caras menos número de cruces después de t lanzamientos.

FIGURA 20.3
Esta tabla de datos mejora la confiabilidad de la simulación registrada en la figura 20.2 pues realiza 1 000 réplicas en lugar de sólo 14.

	I	J	K	L	M
1	**Tabla de datos del juego de lanzamiento**				
2	**de monedas (1 000 réplicas)**				
3					
4			Número		
5		Jugada	de lanzam.	Ganados	
6			5	$3	
7		1	3	$5	
8		2	3	$5	
9		3	7	$1	
10		4	11	–$3	
11		5	13	–$5	
12		6	7	$1	
13		7	3	$5	
14		8	7	$1	
15		9	3	$5	
16		10	9	–$1	
1001		995	5	$3	
1002		996	27	–$19	
1003		997	7	$1	
1004		998	3	$5	
1005		999	9	–$1	
1006		1000	17	–$9	
1007					
1008		Promedio	8.97	–$0.97	

Los *eventos* que cambian el estado del sistema son el lanzamiento de una cara o el de una cruz. El *mecanismo generador de eventos* es la generación de *números aleatorios uniformes* en el intervalo [0, 1], donde

0.0000 a 0.4999 $\Rightarrow$ una cara,

0.5000 a 0.9999 $\Rightarrow$ una cruz.

La *fórmula de transición de estados* consiste en

$$\text{Restablecer } N(t) = \begin{cases} N(t-1)+1 & \text{si el lanzamiento } t \text{ es cara} \\ N(t-1)-1 & \text{si el lanzamiento } t \text{ es cruz.} \end{cases}$$

Entonces, el juego simulado termina en el primer valor de t para el que $N(t) = \pm 3$, donde la *observación* de la muestra que resulta del experimento simulado es $8 - t$, la cantidad ganada (positiva o negativa) en esa jugada.

El siguiente ejemplo ilustra estos bloques de construcción de un modelo de simulación de un importante sistema estocástico de teoría de colas.

EJEMPLO 2 Sistema de colas *M/M/*1

Considere el modelo *M/M/*1 de teoría de colas (entradas Poisson, tiempo de servicio exponencial y un solo servidor) que se presentó al principio de la sección 17.6. Aunque este modelo ya se resolvió en forma analítica, será ilustrativo estudiar la forma como se puede usar simulación. En particular, suponga que los valores de la *tasa de llegadas* λ y la *tasa de servicio* μ son

$$\lambda = 3 \text{ por hora}, \qquad \mu = 5 \text{ por hora}.$$

Para resumir la operación física del sistema, los clientes que llegan se unen a una cola, finalmente son servidos y después se van. Es necesario que el modelo de simulación describa y sincronice la llegada de los clientes y el servicio que se les presta.

Si se inicia en el tiempo 0, el reloj de simulación registra el tiempo (simulado) t que ha transcurrido hasta ahora durante la corrida de simulación. La información sobre el sistema de colas que define el estado actual, es decir, el estado del sistema, es

$N(t)$ = número de clientes en el sistema en el tiempo t.

Los eventos que cambian el estado del sistema son la *llegada* de un cliente o la *terminación del servicio* del cliente que se atiende en este momento (si lo hay). El mecanismo generador de eventos se describirá un poco más adelante. La fórmula de transición de estados es

$$\text{Restablecer } N(t) = \begin{cases} N(t)+1 & \text{si una llegada ocurre en el tiempo } t \\ N(t)-1 & \text{si un servicio se completa en el tiempo } t \end{cases}$$

Existen dos métodos básicos para hacer que avance el reloj simulador y registrar la operación del sistema. No se hizo la distinción entre estos dos métodos en el ejemplo 1 porque coinciden en esa situación tan sencilla. Ahora se describirán e ilustrarán estos dos **métodos para avanzar el tiempo** (incrementos de tiempo fijo e incrementos por evento).

En el mecanismo de **incrementos de tiempo fijo** se usa repetidas veces el siguiente procedimiento de dos pasos.

Resumen de incrementos de tiempo fijo

1. Se *avanza el tiempo* una *cantidad fija* pequeña.
2. Se *actualiza el sistema* determinando cuáles eventos ocurrieron durante este lapso y el estado del sistema que resulta. También se registra la información deseada sobre el comportamiento del sistema.

En el caso del modelo de líneas de espera que se analiza ahora, sólo pueden ocurrir dos tipos de eventos durante cada uno de estos intervalos, esto es, una o más *llegadas* y una o más *terminaciones de servicio*. Es más, la probabilidad de que ocurran dos o más llegadas o dos o más terminaciones de servicio durante una unidad de tiempo es despreciable para este modelo si la unidad de tiempo es relativamente pequeña. En consecuencia, los únicos dos eventos posibles en esa unidad de tiempo que deben investigarse son la llegada de un cliente y la terminación del servicio a uno de ellos. Cada uno de estos eventos tiene una probabilidad conocida.

Con propósitos de ilustración se usará 0.1 de hora (6 minutos) como la cantidad fija pequeña en que avanza el reloj cada vez. (Por lo general, se debe usar un intervalo mucho menor para que en realidad sea despreciable la probabilidad de llegadas o terminaciones de servicio múltiples, pero esta elección crearía más movimiento para propósitos ilustrativos.) Debido a que tanto los tiempos entre llegadas como los de servicio tienen una distribución exponencial, la probabilidad P_A de que un intervalo de 0.1 horas incluya una *llegada* es

$$P_A = 1 - e^{-3/10} = 0.259,$$

y la probabilidad P_D de que incluya una *salida* (terminación de servicio), dado que había un cliente en servicio al inicio del intervalo, es

$$P_D = 1 - e^{-5/10} = 0.393.$$

Para generar de manera aleatoria cualquiera de los dos tipos de eventos de acuerdo con estas probabilidades, el enfoque es similar al del ejemplo 1. De nuevo, se usa la computadora para generar un *número aleatorio uniforme* en el intervalo [0, 1], es decir, una observación aleatoria a partir de una *distribución uniforme* entre 0 y 1. Si este número aleatorio uniforme se denota por r_A,

$$r_A < 0.259 \Rightarrow \text{ocurre una llegada},$$
$$r_A \geq 0.259 \Rightarrow \text{no ocurre una llegada}.$$

De manera similar, con *otro* número aleatorio uniforme r_D,

$$r_D < 0.393 \Rightarrow \text{ocurre una salida},$$
$$r_D \geq 0.393 \Rightarrow \text{no ocurre una salida},$$

dado que había un cliente en servicio al principio del intervalo. Sin clientes en servicio (es decir, sin clientes en el sistema) se supone que no pueden ocurrir salidas durante el intervalo aunque haya ocurrido una llegada.

En la tabla 20.1 se muestran los resultados al usar este enfoque para 10 iteraciones del procedimiento de *incrementos de tiempo fijo* con inicio sin clientes en el sistema y en minutos como unidad de tiempo.

El paso 2 del procedimiento (actualización del sistema) incluye el registro de las medidas de desempeño deseadas del comportamiento agregado del sistema durante este intervalo. Por ejemplo, se puede registrar el *número de clientes* en el sistema de colas y el *tiempo de espera* de cualquier

TABLA 20.1 Aplicación de incrementos de tiempo fijo al ejemplo 2

t, tiempo (min)	*N*(*t*)	r_A	¿Llegada en el intervalo?	r_D	¿Salida en el intervalo?
0	0				
6	1	0.096	Sí	—	
12	1	0.569	No	0.665	No
18	1	0.764	No	0.842	No
24	0	0.492	No	0.224	Sí
30	0	0.950	No	—	
36	0	0.610	No	—	
42	1	0.145	Sí	—	
48	1	0.484	No	0.552	No
54	1	0.350	No	0.590	No
60	0	0.430	No	0.041	Sí

cliente para el que acaba de terminar su espera. Si es suficiente con estimar la media en lugar de la distribución de probabilidad de cada una de estas variables aleatorias, la computadora simplemente añade el valor (si lo hay) al final del intervalo de tiempo presente a una suma acumulativa. Los promedios muestrales se obtienen al terminar la corrida de simulación dividiendo estas sumas acumuladas entre el tamaño de la muestra, es decir, el número total de intervalos y el número total de clientes, respectivamente.

Para ilustrar este procedimiento de estimación, suponga que la corrida de simulación de la tabla 20.1 se usa para estimar W, el tiempo esperado de espera en estado estable de un cliente en el sistema (con el servicio). En esta corrida de simulación llegaron dos clientes, uno en el primer intervalo y el otro en el séptimo, y cada uno permaneció en el sistema durante tres intervalos de tiempo. Entonces, como la duración de cada intervalo es de 0.1 hora, la estimación de W es

$$\text{Est}\{W\} = \frac{3 + 3}{2} \text{ (0.1 hora)} = 0.3 \text{ hora.}$$

Ésta, por supuesto, es sólo una estimación burda, basada en una muestra de sólo dos. (Si se usa la fórmula de W dada en la sección 17.6, el valor verdadero es $W = 1/(\mu - \lambda) = 0.5$ horas.) Por lo general, se usaría una muestra de tamaño mucho más grande.

Otra deficiencia al usar la tabla 20.1 es que esta corrida de simulación se inició sin clientes en el sistema, lo que ocasiona que las observaciones iniciales de los tiempos de espera tiendan a ser un poco menores que el valor esperado cuando el sistema se encuentra estable. Como la meta es estimar el tiempo de espera en *estado estable*, es importante correr la simulación durante un tiempo sin recolectar datos hasta que se piense que el sistema simulado, en esencia, ha alcanzado una condición de estado estable. (En el segundo complemento de este capítulo en el sitio de internet de este libro se describe un método especial para superar este problema.) Este periodo inicial de espera para alcanzar la condición de estado estable antes de recolectar datos se llama **periodo de calentamiento**.

El **incremento por eventos** difiere del incremento de tiempo fijo en que el reloj de simulación se incrementa en una cantidad variable, y no fija, cada vez. Esta cantidad *variable* es el tiempo que transcurre desde que acaba de ocurrir un evento hasta que ocurre el *siguiente evento* de cualquier tipo; esto es, el reloj salta de un evento a otro. A continuación se resume este procedimiento.

Resumen de incremento por eventos

1. Se *avanza el tiempo* hasta el momento en que ocurre el *siguiente evento* de cualquier tipo.
2. Se *actualiza el sistema* al determinar su nuevo estado resultado de este evento y generar, de manera aleatoria (si no se generó antes), el tiempo hasta la siguiente ocurrencia de un evento de cualquier tipo que sea posible desde este estado. También se registra la información deseada sobre el comportamiento del sistema.

En este ejemplo, la computadora debe registrar dos eventos futuros, que son la siguiente llegada y la siguiente terminación de servicio (si se atiende a un cliente). Estos tiempos se obtienen con una observación aleatoria de la distribución de probabilidad de los respectivos tiempos entre llegadas y de servicio. Como antes, la computadora hace esa observación aleatoria al generar y usar un número aleatorio. (Esta técnica se presentará en la sección 20.4.) Por lo tanto, cada vez que ocurra una llegada o una terminación de servicio, la computadora primero determina cuánto tiempo debe transcurrir hasta que ocurra el siguiente evento, suma este tiempo al tiempo actual del reloj y almacena esta suma en un archivo. (Si la terminación de un servicio deja al sistema sin clientes, se pospone la generación del tiempo de la siguiente terminación de servicio hasta que ocurra la siguiente llegada.) Para determinar el siguiente evento, la computadora encuentra el mínimo de los tiempos almacenados en el archivo. Para agilizar los registros, los lenguajes de simulación proporcionan una "rutina de tiempo" que determina el tiempo de ocurrencia y el tipo del siguiente evento, avanza el tiempo y transfiere el control al subprograma apropiado para el tipo de evento.

En la tabla 20.2 se muestran los resultados después de aplicar a este enfoque cinco iteraciones del procedimiento de incrementos por evento; se inicia sin clientes en el sistema y la unidad de tiempo son los minutos. Para su referencia futura, se incluyen los *números aleatorios uniformes* r_A y r_D que se utilizan para generar los tiempos entre llegadas y de servicio, por el método descrito en

TABLA 20.2 Aplicación de incrementos por evento al ejemplo 2

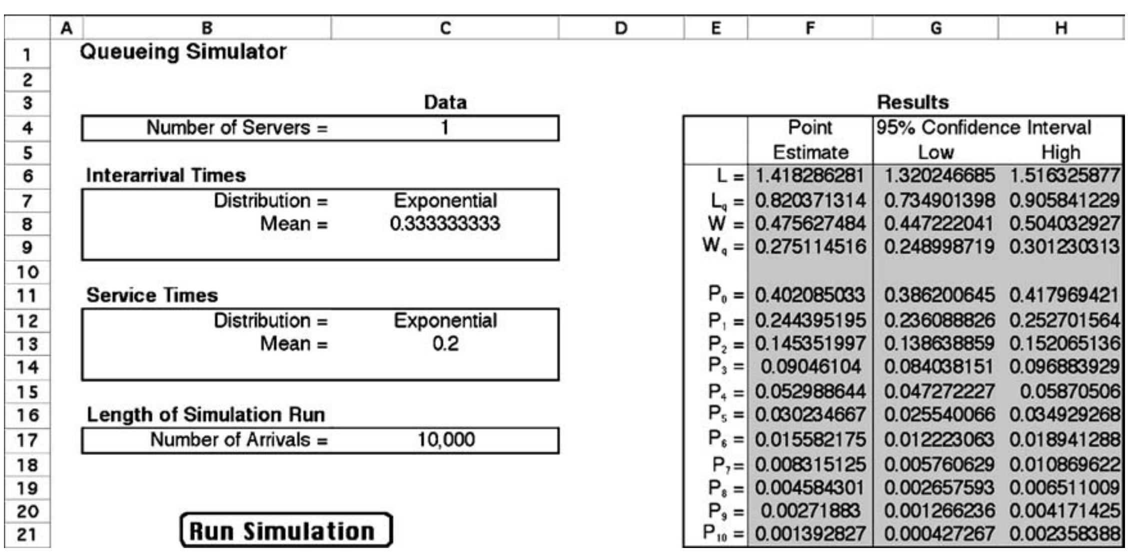
t, tiempo (min)	$N(t)$	r_A	Siguiente tiempo entre llegadas	r_D	Siguiente tiempo de servicio	Siguiente llegada	Siguiente Salida	Siguiente Evento
0	0	0.096	2.019	—	—	2.019	—	Llegada
2.019	1	0.569	16.833	0.665	13.123	18.852	15.142	Salida
15.142	0	—	—	—	—	18.852	—	Llegada
18.852	1	0.764	28.878	0.842	22.142	47.730	40.994	Salida
40.994	0	—	—	—	—	47.730	—	Llegada
47.730	1							

la sección 20.4. Estos r_A y r_D son los mismos que se usan en la tabla 20.1, con el fin de contar con una mejor comparación entre los dos mecanismos de avance del tiempo.

Los archivos de Excel para este capítulo en el OR Courseware incluyen una rutina automática, llamada **simulador de colas** (*Queueing Simulator*), para aplicar el procedimiento de incrementos por evento con varios tipos de sistemas de colas. (Este software es un buen ejemplo se *software de simulación de eventos-discretos* que es ampliamente aplicado en simulación.) El sistema puede tener uno o varios servidores. Se dispone de varias opciones para la distribución de probabilidad (exponencial, Erlang, degenerada, uniforme o exponencial trasladada) de los tiempos entre llegadas y de servicio. En la figura 20.4 se muestra la entrada y salida (en horas) cuando se aplica el simulador al ejemplo actual para una corrida de simulación de 10 000 llegadas de clientes. Según la notación de las diferentes medidas de desempeño de los sistemas de colas de la sección 17.2, la columna *F* da la estimación que proporciona el simulador de cada medida. [Al usar las fórmulas de la sección 17.6 para un simulador de colas *M*/*M*/1, los valores verdaderos de estas medidas son $L = 1.5$, $L_q = 0.9$, $W = 0.5$, $W_q = 0.3$, $P_0 = 0.4$ y $P_n = 0.4(0.6)^n$.] Las columnas *G* y *H* muestran el intervalo de confianza de 95% que corresponde a cada medida. Observe que estos intervalos son un poco más amplios de lo esperado después de una simulación tan larga. En general, se requieren corridas de simulación muy largas para obtener estimaciones relativamente precisas (intervalos de confianza angostos) de las medidas de desempeño de los sistemas de colas (o para la mayoría de los sistemas estocásticos).

El procedimiento de incrementos por evento es mucho más apropiado para este ejemplo y para sistemas estocásticos similares que el de incrementos de tiempo fijo. La técnica de incrementos

FIGURA 20.4
Resultados que se obtuvieron con el simulador de colas en el archivo de Excel de este capítulo para simular el ejemplo 2 durante un periodo de 10 000 llegadas de clientes.



por evento requiere menos iteraciones para cubrir el mismo tiempo simulado y, en lugar de una aproximación, genera un programa preciso de la evolución del sistema.

El procedimiento de incrementos por evento se ilustrará de nuevo en el segundo complemento de este capítulo en el sitio en internet de este libro en el contexto de un experimento estadístico completo para estimar ciertas medidas de desempeño de otro sistema de colas. Esa sección también describe el método estadístico que usa el simulador de colas para obtener estas estimaciones puntuales e intervalos de confianza.

Quedan por contestar varias preguntas pertinentes sobre cómo llevar a cabo un estudio de simulación de este tipo. Las respuestas se presentarán en un contexto más amplio en las secciones subsecuentes.

Más ejemplos en el OR Courseware

Es más sencillo entender los ejemplos de simulación cuando se puede *observar en acción*, en lugar de sólo hablar sobre una página impresa. Por este motivo, el área de simulación del IOR Tutorial incluye una rutina automática llamada "animación de un sistema de colas" (*Animation of a Queueing System*) que muestra una simulación en la que es posible realmente observar a los clientes que entran y salen de un sistema de colas. Por ello, al observar esta animación se ilustra la secuencia de eventos que debería generar el procedimiento de incremento por evento durante la simulación de un sistema de colas. Además, el área de simulación del OR Tutor incluye **dos ejemplos de demostración** que deben verse en este punto.

Ambos ejemplos involucran un banco que planea abrir una nueva sucursal. Las preguntas se refieren a cuántas cajas debe poner en operación y después cuántas cajeras tienen que trabajar en ellas. Por lo tanto, el sistema que se está estudiando es un *sistema de colas*. Sin embargo, al contrario del sistema de colas $M/M/1$ del ejemplo 2 anterior, éste es demasiado complicado para obtener su solución analítica. Tiene múltiples servidores (cajeras), y las distribuciones de probabilidad de los tiempos entre llegadas y de servicio no se ajustan a los modelos estándares de la teoría de colas. Es más, en la segunda demostración se decidió que a cierta clase de clientes (comerciantes) se les debe otorgar una prioridad sin interrupción sobre los otros clientes, pero las distribuciones de probabilidad de estas clases son diferentes. Estas complicaciones son comunes y casi siempre se pueden incorporar con facilidad en un estudio de simulación.

En ambas demostraciones se podrá observar a los clientes que llegan y a los clientes atendidos que salen, al igual que la aplicación del procedimiento de incrementos por evento instrumentados en forma simultánea a la corrida de simulación.

Las demostraciones introducen también una *rutina interactiva* llamada "simulación interactiva de problemas de colas" (*Interactively Simulate Queueing Problem*) en el IOR Tutorial, herramienta que será muy útil para manejar algunos problemas al final de este capítulo.

20.2 ALGUNOS TIPOS COMUNES DE APLICACIONES DE SIMULACIÓN

Debido a su versatilidad, la simulación es una técnica excepcional. Se puede usar (con diferentes grados de dificultad) para investigar virtualmente cualquier tipo de sistema estocástico. Esta versatilidad ha hecho que la simulación sea la técnica de IO que más se usa en estudios que manejan este tipo de sistemas; además, su popularidad continúa en aumento.

Debido a la gran diversidad de aplicaciones, es imposible enumerar todas las áreas específicas en las que se la utiliza. Sin embargo, se hará una descripción breve de algunas categorías importantes de aplicaciones.

Las primeras tres categorías se refieren a tipos de sistemas estocásticos que se consideraron en capítulos anteriores. Es común usar los tipos de modelos matemáticos descritos en esos capítulos para analizar las versiones simplificadas y después aplicar simulación para refinar los resultados.

Diseño y operación de sistemas de colas

La sección 17.3 contiene muchos ejemplos de sistemas de colas comunes, los cuales ilustran la manera en que estos sistemas se extienden a muchas áreas de la sociedad. Se dispone de muchos

modelos matemáticos (incluso los del capítulo 17) para analizar sistemas de colas sencillos. Desafortunadamente, estos modelos sólo pueden proporcionar aproximaciones burdas, en el mejor caso, de sistemas de colas complejos. No obstante, la simulación es apropiada para manejar sistemas de colas muy complicados, y muchas aplicaciones quedan dentro de esta categoría.

Los dos ejemplos de demostración de simulación en el OR Tutor (ambos para decidir el nivel de servicio de cajeros que debe proporcionar un banco) son de este tipo. Debido a que la aplicación de simulación a la teoría de colas es tan común, la rutina automática en el OR Courseware llamada *Queueing Simulator* (ilustrada en la figura 20.4) simula sistemas de colas. (Como se indicó en la sección anterior, esta rutina especial se proporciona en uno de los archivos de Excel de este capítulo.)

Entre las seis aplicaciones ganadoras de premios de los modelos de colas de la sección 17.3, una de ellas usó simulación. Ésta fue una aplicación que involucraba el desarrollo para AT&T de un sistema de basado en PC para ayudar a sus clientes a diseñar o rediseñar sus centros de atención, que produjo ganancias anuales de 750 millones de dólares para los clientes. Esta aplicación de la simulación se describe con más detalle en el Recuadro de aplicación que se presenta en la sección 20.5.

Administración de sistemas de inventarios

En las secciones 18.6 y 18.7 se presentaron modelos para administrar sistemas de inventarios cuando los productos tienen cierta demanda. Sin embargo, los sistemas de inventarios que surgen en la práctica a menudo presentan complicaciones que no se toman en cuenta. Aunque en ocasiones los modelos matemáticos ayudan a analizar esos sistemas complejos, la simulación también suele tener un papel muy importante.

Como ejemplo, un artículo en el número de abril de 1996 de *OR/MS Today* describe un estudio de IO de este tipo que se realizó para *IBM PC Company* en Europa. Al enfrentarse a una fuerte presión por parte de los competidores cada vez más ágiles y agresivos, la compañía debía encontrar una manera de mejorar su desempeño en la entrega de las órdenes. El equipo de IO analizó cómo hacerlo con la simulación de varios rediseños de toda la *cadena de proveedores* de la compañía (la red de instalaciones abarca suministros, manufactura y distribución, e incluye todos los inventarios acumulados en el proceso). Este enfoque condujo a cambios importantes en el diseño y operación de la cadena de proveedores (y de los sistemas de inventario) que mejoraron de manera notable la posición competitiva de la compañía. Se lograron ahorros en costos directos de 40 millones de dólares anuales.

La sección 20.6 ilustrará la aplicación de simulación a un tipo bastante sencillo de sistemas de inventarios.

Estimación de la probabilidad de terminar un proyecto a tiempo

Una de las preocupaciones más grandes de un administrador de proyecto es si su equipo será capaz de terminarlo en la fecha convenida de entrega. La sección 22.4 (en el sitio en internet de este libro) describe cómo se puede usar el enfoque de PERT de tres estimaciones para obtener una estimación burda de la probabilidad de cumplir con la fecha de entrega con el plan de proyecto actual. En esa sección también se describen tres aproximaciones simplificadas de este enfoque, que permiten estimar esta probabilidad. Desafortunadamente, debido a estas aproximaciones, las estimaciones que se obtienen siempre son optimistas, y algunas veces demasiado.

En consecuencia, cada vez es más común usar simulación para obtener mejores estimaciones de esta probabilidad. Esto incluye la generación de observaciones aleatorias a partir de distribuciones de probabilidad para la duración de las diferentes actividades en los proyectos. Si se usa la red del proyecto, entonces directamente se simula cuándo comienza cada actividad y cuándo termina y, de esta manera, también se simula cuándo termina el proyecto. La repetición de esta simulación algunas miles de veces (en una corrida de computadora) puede proporcionar una estimación muy buena de la probabilidad de cumplir con la fecha de entrega.

Este tipo específico de aplicación se ilustra en la sección 28.2 del sitio en internet de este libro.

Recuadro de aplicación

Desde su fundación en 1914, **Merrill Lynch** ha sido una firma líder en servicios financieros que se esfuerza por convertir Wall Street en Main Street, esto es, poner los mercados financieros al alcance de todo el mundo. Emplea una fuerza de ventas muy capacitada formada por más de 15 000 asesores financieros en Estados Unidos y tiene presencia en 36 países. Como una de las compañías de *Fortune 100* con ganancias netas de 26 billones (10^9) de dólares en 2005, tiene bajo su control activos de los clientes que suman más de 1.7 trillones (10^{12}) de dólares.

Al enfrentarse con una competencia cada vez mayor de firmas de correduría de descuento y de correduría electrónica, a finales de 1998 se decidió formar un grupo de tarea con el fin de recomendar un producto o servicio para satisfacer las exigencias del mercado. Al grupo de IO de Merrill Lynch se le encomendó hacer un análisis detallado de dos nuevas opciones de precios para los clientes. Una de ellas debería reemplazar el cargo por transacciones de forma individual mediante el cargo de un porcentaje fijo de los activos de los clientes y, después, permitir un número ilimitado de transacciones libres y acceso total a un asesor financiero. La otra opción permitiría que los clientes invirtieran en línea directamente con una cuota fija reducida por transacción sin consultar a un asesor de inversiones.

El gran reto que encaró el grupo de IO fue determinar el "punto sano" para los precios de estas opciones que fuera capaz de aumentar los negocios de la firma e incrementar sus ganancias y a la vez minimizar el riesgo de perderlas. Una herramienta clave para abordar este problema fue el uso de la *simulación*. Para llevar a cabo un *estudio de simulación mayor*, el grupo formó y evaluó un enorme volumen de datos acerca de los activos y de la actividad comercial de los cinco millones de clientes de la firma. En cada segmento de la base de clientes se hizo un análisis minucioso relativo a su comportamiento de aceptar ofertas mediante el uso del criterio administrativo, la investigación de mercados y la experiencia con los clientes. Con toda esta información como insumo, el grupo procedió a formular y a correr un *modelo de simulación* con varios escenarios de precios para poder identificar el punto sano de éstos.

La implementación de estos resultados tuvo un profundo efecto en la posición competitiva de Merrill Lynch, colocándola en una posición de liderazgo en la industria. En lugar de seguir perdiendo terreno en la feroz competencia que enfrentaba, *los activos de los clientes administrados por la compañía aumentaron* **22 billones de dólares** (10^9) *y su ganancia alcanzó los* **80 millones de dólares** en *un periodo de 18 meses*. El presidente ejecutivo de Merrill Lynch llamó a esta nueva estrategia "la decisión más importante que hemos tomado, como firma (en los últimos 20 años)". Esta aplicación de la simulación representó un enorme éxito y llevó a Merrill Lynch a obtener el primer lugar en la competencia internacional por el Premio Franz Edelman al desempeño en investigación de operaciones y ciencias de la administración.

Fuente: S. Altschuler, D. Batavia, J. Bennett, R. Labe, B. Liao, R. Nigam y J. Oh: "Pricing Analysis for Merrill Lynch Integrated Choice", en *Interfaces*, **32**(1): 5-19, enero-febrero de 2002. (En el sitio en internet de este libro, www.mhhe.com/hillier, se proporciona una liga hacia este artículo).

Diseño y operación de sistemas de manufactura

Los estudios muestran de manera firme que una gran proporción de aplicaciones de simulación incluyen sistemas de manufactura. Muchos de ellos se pueden considerar como un sistema de colas de algún tipo (como un sistema de colas donde las máquinas son los servidores y los trabajos que se deben procesar son los clientes). Sin embargo, diversas complicaciones inherentes a estos sistemas (descomposturas ocasionales de las máquinas, artículos defectuosos que necesitan retrabajo y tipos múltiples de tareas) están fuera del alcance de los modelos de colas usuales. Esas complicaciones no son un problema para la simulación.

Algunos ejemplos de los tipos de preguntas que pueden hacerse son:

1. ¿Cuántas máquinas de cada tipo debe haber?
2. ¿Cuántas unidades de manejo de materiales de cada tipo debe haber?
3. Si se consideran las fechas de terminación del proceso completo de producción, ¿qué regla debe usarse para elegir el orden en que debe procesarse los trabajos que están ahora en una máquina?
4. ¿Cuáles son las fechas de terminación realistas de los trabajos?
5. ¿Cuáles serán las operaciones "cuello de botella" en un nuevo proceso de producción de acuerdo con el diseño actual?
6. ¿Cuál será la salida (tasa de producción) de un nuevo proceso de producción?

La referencia seleccionada A1 describe una aplicación de este último tipo que ha sido ganadora de premios. General Motors Corporation tuvo tanto éxito al aplicar la simulación para predecir y mejorar la capacidad de sus líneas de producción, que pudo incrementar sus ganancias y ahorrar más de 2 100 millones de dólares en 30 plantas de vehículos en 10 países.

Diseño y operación de sistemas de distribución

Cualquier corporación de manufactura grande necesita un *sistema de distribución* eficiente para sus productos, desde las fábricas y almacenes hasta los clientes. Existen muchas incertidumbres relacionadas con la operación de esos sistemas. ¿Cuándo estarán disponibles los vehículos para enviar los productos? ¿Cuánto tiempo toma un envío? ¿Cuál será la demanda por parte de los diferentes clientes? Al generar observaciones aleatorias a partir de una distribución de probabilidad relevante, las simulaciones pueden manejar con facilidad estos tipos de incertidumbre. Por ello, con frecuencia se usan para probar diferentes posibilidades para mejorar el diseño y operación de estos sistemas.

Una aplicación de este tipo ganadora de un premio se describe en el número de enero-febrero de 1991 de *Interfaces*. La *Reynolds Metal Company* gasta más de 250 millones de dólares anuales para entregar sus productos y recibir la materia prima. Los embarques se hacen por camión, ferrocarril y avión por medio de una red de más de cien puntos de envío que incluyen plantas, almacenes y proveedores. Ante ello, se aplicó una combinación de programación entera binaria mixta (capítulo 11) y simulación para diseñar un nuevo sistema de distribución con despacho central. El nuevo sistema mejoró el tiempo de entrega de los embarques y redujo los costos de fletes anuales en más de 7 millones de dólares.

Análisis de riesgo financiero

El análisis de riesgo financiero fue una de las primeras áreas de aplicación de simulación y continúa muy activa. Por ejemplo, considere la evaluación de una propuesta de inversión de capital con flujos de efectivo inciertos. Al generar observaciones aleatorias a partir de distribuciones de probabilidad de los flujos de efectivo en cada uno de los periodos (y considerar las relaciones entre los periodos), la simulación puede generar miles de escenarios acerca del resultado de esa inversión. Este procedimiento proporciona una *distribución de probabilidad* del rendimiento (como por ejemplo un valor presente neto) de la inversión. Esta distribución (que suele llamarse *perfil de riesgo*) permite a la administración evaluar el riesgo involucrado al hacer esa inversión.

Un enfoque similar permite analizar el riesgo asociado con la inversión en diferentes valores, inclusive instrumentos financieros menos comunes como depósitos, requerimientos, futuros, acciones, opciones, etcétera.

La sección 28.4 del sitio en internet de este libro incluye un ejemplo de simulación para el análisis de riesgo financiero.

Aplicaciones en el cuidado de la salud

El cuidado de la salud es otra área donde, igual que en la evaluación de inversiones riesgosas, el análisis de la incertidumbre futura es el eje de la toma de decisiones. Sin embargo, en lugar de lidiar con flujos de efectivo futuros inciertos, las incertidumbres ahora se refieren a la evolución de enfermedades humanas.

Los siguientes son algunos ejemplos de los tipos de simulaciones en computadora que se han realizado para ayudar en el diseño de los sistemas del cuidado de la salud.

1. Simulación del uso de los recursos de un hospital para tratar a pacientes con problemas coronarios.
2. Simulación de los gastos médicos bajo planes de seguros alternativos.
3. Simulación del costo y eficacia de la detección temprana de enfermedades.
4. Simulación del uso de servicios quirúrgicos complejos en un centro médico.
5. Simulación del tiempo y lugar de llamadas que solicitan servicios de ambulancia.
6. Simulación de la compatibilidad de riñones donados con receptores de trasplante.
7. Simulación de la operación de una sala de urgencias.

Aplicaciones en otras industrias de servicios

Igual que el cuidado de la salud, otras industrias de servicio son campos fértiles para aplicar la simulación. Estas industrias incluyen servicios del gobierno, bancos, administración de hoteles, restaurantes, instituciones educativas, planeación de desastres, las fuerzas armadas, parques de diversión y otras más. En muchos casos, en realidad, los sistemas simulados son sistemas de colas de algún tipo.

En la referencia seleccionada A5 se describe una aplicación ganadora de un premio en esta categoría. El *servicio postal de Estados Unidos* identificó la *tecnología de automatización* como la única manera de poder manejar el creciente volumen de correo al mismo tiempo que conservar un precio competitivo y satisfacer las metas de servicio. Se requirió una planeación extensa durante varios años para la conversión a un sistema automatizado en gran parte para cumplir con estas metas. La base del análisis que condujo al plan adoptado se realizó con un modelo de simulación integral llamado "modelo para la evaluación de alternativas de tecnología" o META (*model for evaluating technology alternatives*). Primero se aplicó el modelo global en todo el país y después se redujo a la entidad para llevar a cabo una planeación más detallada. El plan resultante requirió una inversión de capital acumulada de 12 000 millones de dólares, pero también se proyectó para lograr ahorros en mano de obra de 4 000 millones al año. Otra consecuencia de esta exitosa aplicación de simulación fue que se reconoció el valor de las herramientas de IO en los más altos niveles del servicio postal de Estados Unidos. Las técnicas de IO continúan en uso en el departamento de planeación, tanto en las oficinas generales como en las divisiones.

Aplicaciones militares

Probablemente no exista otro sector de la sociedad donde la simulación se utilice tan extensamente como el militar. En realidad, la dependencia de las fuerzas armadas de la simulación para llevar a cabo juegos de guerra se remonta a varios siglos atrás y las diferentes academias militares de Estados Unidos la han incluido en sus planes de estudio desde que esta rama se inició. Sin embargo, el advenimiento de poderosas computadoras ha sido un factor muy importante para el crecimiento exponencial del uso de la simulación en el ambiente militar, en especial por parte del Departamento de Defensa. Hoy en día, el uso del juego de la guerra con el fin de simular operaciones militares se utiliza de forma rutinaria para planear operaciones militares, para actualizar la doctrina castrense y para entrenar a futuros oficiales. La simulación también resulta ser una herramienta muy útil para tomar decisiones militares acertadas.

Nuevas aplicaciones

Cada año se realizan más aplicaciones innovadoras de simulación. Muchas de ellas se anuncian en la reunión Winter Simulation Conference, que se realiza cada diciembre en alguna ciudad de Estados Unidos. Desde sus inicios en 1967, ha sido una institución precursora en el campo de la simulación. Asisten cerca de mil participantes, académicos y practicantes por igual. En ella se presentan cientos de artículos acerca de avances metodológicos y aplicaciones innovadoras.

20.3 GENERACIÓN DE NÚMEROS ALEATORIOS

Como se demostró en los ejemplos de la sección 20.1, cuando se pone en práctica un modelo de simulación, son necesarios números aleatorios para obtener observaciones aleatorias a partir de distribuciones de probabilidad. Un método para generar estos números aleatorios consiste en utilizar un dispositivo físico como un disco que da vueltas o un aleatorizador electrónico. De esta manera se han generado varias tablas de números aleatorios, entre las que se cuenta una que contiene un millón de dígitos aleatorios, publicada por la Rand Corporation. En la tabla 20.3 se presenta un extracto de ella.

En la actualidad, los dispositivos físicos han sido sustituidos por computadoras como la fuente principal de generadores de números aleatorios. Por ejemplo, en la sección 20.1 se señaló que Excel usa la función ALEATORIO() para alcanzar este objetivo. Muchos otros paquetes de software cuentan con la generación de números aleatorios cuando los requiere la corrida de simulación.

Características de los números aleatorios

El procedimiento que aplica una computadora para obtener números aleatorios se llama *generador de números aleatorios*.

Un **generador de números aleatorios** es un algoritmo que produce secuencias de números que siguen una distribución de probabilidad específica y tienen la apariencia de aleatoriedad.

TABLA 20.3 Tabla de dígitos aleatorios

09656	96657	64842	49222	49506	10145	48455	23505	90430	04180
24712	55799	60857	73479	33581	17360	30406	05842	72044	90764
07202	96341	23699	76171	79126	04512	15426	15980	88898	06358
84575	46820	54083	43918	46989	05379	70682	43081	66171	38942
38144	87037	46626	70529	27918	34191	98668	33482	43998	75733
48048	56349	01986	29814	69800	91609	65374	22928	09704	59343
41936	58566	31276	19952	01352	18834	99596	09302	20087	19063
73391	94006	03822	81845	76158	41352	40596	14325	27020	17546
57580	08954	73554	28698	29022	11568	35668	59906	39557	27217
92646	41113	91411	56215	69302	86419	61224	41936	56939	27816
07118	12707	35622	81485	73354	49800	60805	05648	28898	60933
57842	57831	24130	75408	83784	64307	91620	40810	06539	70387
65078	44981	81009	33697	98324	46928	34198	96032	98426	77488
04294	96120	67629	55265	26248	40602	25566	12520	89785	93932
48381	06807	43775	09708	73199	53406	02910	83292	59249	18597
00459	62045	19249	67095	22752	24636	16965	91836	00582	46721
38824	81681	33323	64086	55970	04849	24819	20749	51711	86173
91465	22232	02907	01050	07121	53536	71070	26916	47620	01619
50874	00807	77751	73952	03073	69063	16894	85570	81746	07568
26644	75871	15618	50310	72610	66205	82640	86205	73453	90232

Fuente: Reproducido con permiso de The Rand Corporation: *A Million Random Digits with 100,000 Normal Deviates.* Copyright, The Free Press, Glencoe, IL, 1955, parte superior p. 182.

La referencia a *secuencias de números aleatorios* significa que el algoritmo produce muchos números aleatorios en serie. Aunque un usuario en particular puede requerir sólo unos cuantos, en general el algoritmo debe ser capaz de producir muchos. La *distribución de probabilidad* implica que se puede asociar una afirmación probabilística con la ocurrencia de cada número producido por el algoritmo.

Se reservará el término **número aleatorio** para hablar de una observación aleatoria a partir de alguna *distribución uniforme*, de manera que todos los números posibles son *igualmente probables*. Cuando el interés se centra en alguna otra distribución de probabilidad (como en la siguiente subsección), se hablará de *observaciones aleatorias* a partir de esa distribución.

Los números aleatorios se pueden dividir en dos categorías principales: números aleatorios enteros y números aleatorios uniformes, definidos de la siguiente manera:

> Un **número aleatorio entero** es una observación aleatoria de una *distribución uniforme discretizada* en el intervalo $\underline{n}, \underline{n}+1, \ldots, \overline{n}$. Las probabilidades de esta distribución son
>
> $$P(\underline{n}) = P(\underline{n}+1) = \cdots = P(\overline{n}) = \frac{1}{\overline{n} - \underline{n} + 1}.$$
>
> Por lo general, $\underline{n} = 0$ o 1 y éstos son valores convenientes para la mayoría de las aplicaciones. (Si $\underline{n}$ tiene otro valor, entonces al restar ya sea $\underline{n}$ o bien $\underline{n} - 1$ del número aleatorio entero cambia el límite inferior del intervalo a 0 o a 1.)
>
> Un **número aleatorio uniforme** es una observación aleatoria a partir de una *distribución uniforme* (continua) en un intervalo $[a, b]$. La función de densidad de probabilidad de esta distribución uniforme es
>
> $$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } a \le x \le b \\ 0 & \text{de otra manera.} \end{cases}$$
>
> Cuando a y b no se especifican, se supone que $a = 0$ y $b = 1$.

Los números aleatorios generados en un inicio por una computadora casi siempre son números aleatorios enteros. Sin embargo, si se desea, es posible convertirlos en números aleatorios uniformes de la manera siguiente:

> Para un *número aleatorio entero* dado entre 0 y $\bar{n}$, al dividir este número entre $\bar{n}$ se obtiene (aproximadamente) un *número aleatorio uniforme*. (Si $\bar{n}$ es pequeño, esta aproximación debe mejorarse al sumar $\frac{1}{2}$ al número aleatorio entero para después dividir entre $\bar{n} + 1$.)

Éste es el método común que se usa para generar números aleatorios uniformes. Cuando se usan valores muy grandes de $\bar{n}$, en esencia es un método exacto.

En sentido estricto, los números generados por una computadora no se deben llamar números aleatorios porque son predecibles y se pueden reproducir (lo que a veces es una ventaja), dado el número aleatorio generador que se use. Por ello, en ocasiones se les llama **números pseudoaleatorios**. No obstante, el punto importante es que, en forma satisfactoria, hacen las veces de números aleatorios en la simulación si el método que se usa para generarlos es válido.

Se han propuesto varios procedimientos estadísticos bastante complejos para probar si una sucesión de números generada tiene una apariencia de aleatoriedad aceptable. En esencia, los requisitos son que cada número sucesivo tenga una probabilidad igual de tomar cualquiera de los valores posibles y que sea estadísticamente independiente de los otros números en la sucesión.

Métodos congruenciales para generar números aleatorios

Se cuenta con varios generadores de números aleatorios, de los cuales los más populares son los *métodos congruenciales* (aditivo, multiplicativo y mixto). El método congruencial mixto incluye características de los otros dos, por lo que se presentará en primer lugar.

El **método congruencial mixto** genera una *sucesión* de números aleatorios enteros en un intervalo de 0 a $m - 1$. Este método siempre calcula el siguiente número a partir del último que obtuvo, dado un número aleatorio inicial x_0, llamado **semilla**, que se puede obtener de alguna fuente publicada como la tabla Rand. En particular, calcula el $(n + 1)$-ésimo número aleatorio x_{n+1} a partir del n-ésimo número aleatorio x_n con la relación de recurrencia

$$x_{n+1} \equiv (ax_n + c)(\text{módulo } m),$$

donde a, c y m son enteros positivos ($a < m$, $c < m$). Esta notación matemática significa que x_{n+1} es el *residuo* cuando $ax_n + c$ se divide entre m. En consecuencia, los valores *posibles* de x_{n+1} son $0, 1, \ldots, m - 1$, de manera que m representa el número deseado de valores *diferentes* que se puede generar como números aleatorios.

A manera de ilustración, suponga que $m = 8$, $a = 5$, $c = 7$ y $x_0 = 4$. En la tabla 20.4 se calculó la sucesión de números aleatorios que se obtuvo (esta sucesión no se puede continuar, puesto que sólo se repetirían los números en el mismo orden). Observe que esta sucesión incluye los ocho números posibles sólo una vez. Esta propiedad es necesaria para una sucesión de números *aleatorios* enteros, pero no ocurre con algunos valores de a y c (intente $a = 4$, $c = 7$ y $x_0 = 3$). Por fortuna, existen reglas para elegir los valores de a y c que garantizan esta propiedad (no hay restricciones sobre la semilla, x_0, porque sólo afecta a la sucesión en el punto en el que comienza y no en la progresión de los números).

La cantidad de números consecutivos de una sucesión antes de que se repita se conoce como **longitud de ciclo**. En consecuencia, la longitud de ciclo en el ejemplo es 8. La longitud de ciclo *máxima* es m, de manera que sólo los valores de a y c que se consideran son los que conducen a esta longitud de ciclo máxima.

En la tabla 20.5 se ilustra la conversión de números aleatorios en números aleatorios uniformes. La columna de la izquierda proporciona los números aleatorios enteros que se obtuvo en la última columna de la tabla 20.4. La última columna proporciona los números aleatorios uniformes correspondientes a partir de la fórmula

$$\text{Número aleatorio uniforme} = \frac{\text{número aleatorio entero} + \dfrac{1}{2}}{m}.$$

TABLA 20.4 Ilustración del método congruencial mixto

n	x_n	$5x_n + 7$	$(5x_n + 7)/8$	x_{n+1}
0	4	27	$3 + \frac{3}{8}$	3
1	3	22	$2 + \frac{6}{8}$	6
2	6	37	$4 + \frac{5}{8}$	5
3	5	32	$4 + \frac{0}{8}$	0
4	0	7	$0 + \frac{7}{8}$	7
5	7	42	$5 + \frac{2}{8}$	2
6	2	17	$2 + \frac{1}{8}$	1
7	1	12	$1 + \frac{4}{8}$	4

Observe que cada número aleatorio uniforme se encuentra en el punto medio de uno de los ocho intervalos de 0 a 0.125, de 0.125 a 0.25, . . . , de 0.875 a 1. El valor tan pequeño de $m = 8$ no permite obtener otros valores en el intervalo [0, 1], por lo que se obtiene aproximaciones bastante burdas de los números aleatorios uniformes. En la práctica, casi siempre se usan valores *mucho* más grandes.

La sección Worked Examples del sitio en internet de este libro incluye **otro ejemplo** de aplicación del método congruencial mixto con un valor de m relativamente pequeño ($m = 16$), para después convertir los números aleatorios enteros que resultan en números aleatorios uniformes. Más adelante, este ejemplo explora los problemas que surgen por el uso de un valor tan pequeño de m.

Cuando se trabaja con una computadora binaria con tamaño de palabra de b bits, la elección normal de m es $m = 2^b$; éste es el número total de enteros no negativos que se pueden expresar dentro de la capacidad del tamaño de la palabra. (Cualquier entero no deseado que aparezca en la sucesión de números aleatorios, simplemente no se usa.) Con esta elección de m se puede asegurar que cada número posible ocurre exactamente una vez antes de que se repita cualquier número si se seleccionan $a = 1, 5, 9, 13, \ldots$, y $c = 1, 3, 5, 7, \ldots$ Si se trata de una computadora decimal con un tamaño de palabra de d dígitos, la elección normal de m es 10^d y se garantiza la misma propiedad si se eligen valores de $a = 1, 21, 41, 61, \ldots$, y $c = 1, 3, 7, 9, 11, 13, 17, 19, \ldots$ (es decir, todos los enteros positivos *impares excepto* aquellos que terminan en 5). Se puede hacer la elección específica con base en la *correlación serial* entre los números sucesivos generados, la cual difiere mucho entre las distintas opciones.[1]

TABLA 20.5 Conversión de números aleatorios enteros en números aleatorios uniformes

Número aleatorio entero	Número aleatorio uniforme
3	0.4375
6	0.8125
5	0.6875
0	0.0625
7	0.9375
2	0.3125
1	0.1875
4	0.5625

[1] Vea R. R. Coveyou, "Serial Correlation in the Generation of Pseudo-Random Numbers", en *Journal of the Association of Computing Machinery*, **7**: 72-74, 1960.

En ocasiones se desea generar números aleatorios enteros que sólo tengan un número reducido de dígitos. Por ejemplo, suponga que sólo se requieren tres dígitos, de manera que los valores posibles se expresan como 000, 001, . . . , 999. En tal caso, el procedimiento normal todavía usa $m = 2^b$ o $m = 10^d$, con lo que se puede generar una gran cantidad de números aleatorios antes de que la sucesión se repita. No obstante, con excepción del cálculo del siguiente número aleatorio entero de esta sucesión, es posible descartar todos los dígitos de cada número generado excepto tres para obtener los números enteros aleatorios de tres dígitos que se buscan. Una convención es tomar los últimos tres dígitos (es decir, los tres dígitos de la cola).

El **método congruencial multiplicativo** corresponde al caso especial del método congruencial mixto en el que $c = 0$. El **método congruencial aditivo** también es parecido, pero establece $a = 1$ y sustituye c por algún número aleatorio anterior a x_n en la sucesión, por ejemplo, x_{n-1} (así, se requiere más de una semilla para iniciar el cálculo de la sucesión).

El método congruencial mixto proporciona una gran flexibilidad para elegir un generador de números aleatorios particular (una combinación específica de a, c y m). Sin embargo, se requiere tener mucho cuidado al seleccionar el generador de números aleatorios porque la mayoría de las combinaciones de valores de a, c y m conducen a propiedades indeseables (por ejemplo, una longitud de ciclo menor a m). Cuando los investigadores identifican generadores de números aleatorios atractivos realizan pruebas extensas para encontrar alguna falla, lo cual puede conducir a un mejor generador. Por ejemplo, hace algunos años, $m = 2^{31}$ fue considerada una opción atractiva, pero ahora los expertos la consideran inaceptable y recomiendan en su lugar el uso de ciertos números mucho más grandes, incluso valores específicos de m cercanos a 2^{191}.[2]

20.4 GENERACIÓN DE OBSERVACIONES ALEATORIAS A PARTIR DE UNA DISTRIBUCIÓN DE PROBABILIDAD

Dada una sucesión de números aleatorios enteros, cabe preguntarse cómo se puede generar una sucesión de observaciones aleatorias que sigan una distribución de probabilidad dada. Se dispone de varios enfoques, según la naturaleza de la distribución.

Distribuciones discretas sencillas

En el caso de algunas distribuciones discretas sencillas se puede usar una secuencia de números aleatorios *enteros* para generar observaciones aleatorias en forma directa. Sólo se asignan los valores posibles de un número aleatorio a los diferentes resultados de la distribución de probabilidad en proporción directa a las respectivas probabilidades de esos resultados.

Por ejemplo, en la sección 20.1, donde se simula lanzar una moneda, los resultados posibles son cara o cruz, donde cada uno tiene probabilidad de $\frac{1}{2}$. Entonces, en lugar de usar números aleatorios uniformes (como se hizo en la sección 20.1), sería suficiente con usar *dígitos aleatorios* para generar los resultados. Cinco de los diez valores posibles de un dígito (por ejemplo, 0, 1, 2, 3, 4) se asignarían a cara y los otros cinco (5, 6, 7, 8, 9) a cruz.

Como otro ejemplo, considere la distribución de probabilidad del resultado de tirar dos dados. Se sabe que la probabilidad de obtener 2 es $\frac{1}{36}$ (igual que la probabilidad de obtener 12), que la probabilidad de obtener 3 es $\frac{2}{36}$, etc. Entonces, se debe asociar $\frac{1}{36}$ de los valores posibles de un número aleatorio entero al hecho de obtener un 2, $\frac{2}{36}$ de los valores al 3, etc. Así, si se usan números aleatorios enteros de dos dígitos, se elige 72 de los 100 valores, y se rechaza el número aleatorio entero si toma cualquiera de los otros 28 valores. Entonces, 2 de los 72 valores posibles (por ejemplo, 00 y 01) se asocian al resultado 2, cuatro de ellos (por ejemplo, 02, 03, 04 y 05) al resultado 3, y así sucesivamente.

Utilizar números aleatorios *enteros* de este modo es conveniente si se obtienen de una tabla de números aleatorios o se generan directamente a través de un método congruencial. Sin embargo, si se realiza una simulación en computadora, casi siempre es más conveniente que ésta genere números aleatorios *uniformes* para después usarlos en la forma correspondiente. Todos los métodos subsecuentes para generar observaciones aleatorias usan números aleatorios uniformes.

[2] Para recomendaciones sobre la elección de un generador de números aleatorios, vea P. L'Ecuyer, R. Simard, E. J. Chen y W. D. Kelton, "An Object-Oriented Random-Number Package with Many Long Streams and Substreams", en *Operations Research*, **50:** 1073-1075, 2002.

Método de transformación inversa

Cuando se trata de distribuciones más complicadas, ya sean discretas o continuas, a veces se puede usar una generalización de este enfoque llamado *método de transformación inversa* para generar observaciones aleatorias. Sea X la variable aleatoria involucrada y sea su función de distribución acumulada la siguiente

$$F(x) = P\{X \leq x\}.$$

En consecuencia, la generación de cada observación requiere los siguientes dos pasos.

Resumen del método de transformada inversa

1. Genere un *número aleatorio uniforme* r entre 0 y 1.
2. Establezca $F(x) = r$ y despeje x, que es entonces la observación aleatoria deseada que sigue la distribución de probabilidad dada.

Este procedimiento se ilustra en la figura 20.5 para el caso en que se grafica la función de distribución acumulada $F(x)$, y ocurre que el número aleatorio uniforme r es 0.5269.

Aunque el procedimiento gráfico que se ilustra en la figura 20.5 es útil para llevar a cabo una simulación manual, la computadora debe disponer de algún enfoque alternativo. En el caso de las distribuciones *discretas*, se puede usar un *enfoque con tabla de búsqueda* si se construye una tabla que proporciona un "intervalo" (salto) en el valor de $F(x)$ para cada valor posible de $X = x$. Excel proporciona la función VLOOKUP para poner en práctica este enfoque al realizar una simulación en hoja de cálculo.

Para ilustrar cómo trabaja esta función, suponga que una compañía simula el *programa de mantenimiento* de sus máquinas. El tiempo entre descomposturas de una de ellas es siempre 4, 5 o 6 días, donde estos tiempos ocurren con probabilidades respectivas de 0.25, 0.5 y 0.25. El primer paso para simular estas descomposturas es crear la tabla que se muestra en la figura 20.6 en algún lugar de la hoja de cálculo. Observe que los números de la segunda columna proporcionan la probabilidad acumulada *previa* al número de días en la tercera columna. La segunda y tercera columnas (sin encabezados) constituyen la "tabla de búsqueda". La función VLOOKUP tiene tres argumentos. El primero da la dirección de la celda que contiene el número aleatorio uniforme que está en uso. El segundo identifica las direcciones del intervalo de celdas de la tabla de búsqueda. El tercero indica qué columna de la tabla (la segunda y tercera columnas de la figura 20.6) proporciona la observación aleatoria, por lo que este argumento es igual a 2 en este caso. La función VLOOKUP con estos tres argumentos se introduce como la ecuación de cada celda de la hoja de cálculo donde debe haber una observación aleatoria que sigue la distribución.

En el caso de ciertas distribuciones *continuas* se puede usar el método de transformación inversa en la computadora si primero se despeja *x analíticamente* de la ecuación $F(x) = r$. En la

FIGURA 20.5
Ilustración del método de la transformada inversa para obtener una observación aleatoria a partir de una distribución de probabilidad dada.

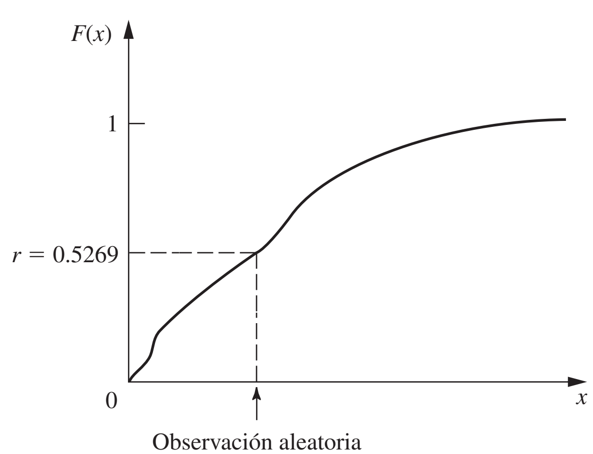


FIGURA 20.6
La tabla debe construirse en una hoja de cálculo para usar la función BUSCARV de Excel para implantar el método de la transformación inversa en el ejemplo del programa de mantenimiento.

Distribución de tiempo entre descomposturas

Probabilidad	Acumulada	Número de días
0.25	0	4
0.5	0.25	5
0.25	0.75	6

sección Worked Examples del sitio en internet de este libro se incluye **un ejemplo** que ilustra este enfoque (después de aplicar el enfoque gráfico).

Este enfoque también se ilustrará con la distribución exponencial.

Distribuciones exponencial y Erlang

Como se mencionó en la sección 17.4, la función de distribución acumulada de la **distribución exponencial** está dada por

$$F(x) = 1 - e^{-\alpha x}, \qquad \text{para } x \geq 0,$$

donde $1/\alpha$ es la media de la distribución. Si se iguala $F(x) = r$ se obtiene

$$1 - e^{-\alpha x} = r,$$

de manera que

$$e^{-\alpha x} = 1 - r.$$

Por lo tanto, si se aplica el logaritmo natural en ambos lados,

$$\ln e^{-\alpha x} = \ln (1 - r),$$

por lo que

$$-\alpha x = \ln (1 - r),$$

que conduce a

$$x = \frac{\ln (1 - r)}{-\alpha}$$

Ahora observe que $1 - r$ es en sí mismo un número aleatorio uniforme. Por lo tanto, para evitar una resta, es una práctica muy común usar sólo el número aleatorio uniforme *original* r directamente en lugar de $1 - r$, lo cual da

$$\text{Observación aleatoria} = \frac{\ln r}{-\alpha}$$

que es la observación que se busca a partir de la distribución exponencial.

Esta aplicación directa del método de transformación inversa proporciona la forma más sencilla de generar observaciones aleatorias a partir de una distribución exponencial. (Otras técnicas más complejas, que pueden ser más rápidas con computadora que calcular el logaritmo, han sido desarrolladas para esta distribución.)[3]

Una extensión natural de este procedimiento de la distribución exponencial se puede usar para generar una observación aleatoria a partir de una **distribución Erlang** (o gamma) (vea la sección 17.7). La suma de k variables aleatorias exponenciales independientes, cada una con media $1/(k\alpha)$ tiene distribución Erlang con parámetro de forma k y media $1/\alpha$. Entonces, dada una sucesión de

[3] Por ejemplo, vea J. H. Ahrens y V. Dieter, "Efficient Table-Free Sampling Methods for Exponencial, Cauchy, and Normal Distributions", en *Communications of the ACM*, **31**: 1330-1337, 1988.

k números aleatorios entre 0 y 1, como $r_1, r_2, \ldots, r_k$, la observación aleatoria deseada a partir de la distribución Erlang es

$$x = \sum_{i=1}^{k} \frac{\ln r_i}{-k\alpha},$$

que se reduce a

$$x = -\frac{1}{k\alpha} \ln \left[\prod_{i=1}^{k} r_i \right],$$

donde Π denota multiplicación.

Distribuciones normal y ji cuadrada

Una técnica en particular sencilla para generar observaciones aleatorias a partir de una **distribución normal** se obtiene al aplicar el *teorema del límite central*. Debido a que un número aleatorio uniforme tiene una *distribución uniforme* entre 0 y 1, resulta que tiene media de 1/2 y desviación estándar $1/\sqrt{12}$. En consecuencia, este teorema implica que la suma de n números aleatorios uniformes tiene una distribución normal aproximada con media $n/2$ y desviación estándar $\sqrt{n/12}$. Por lo tanto, si $r_1, r_2, \ldots, r_n$ constituyen una muestra de números aleatorios uniformes, entonces

$$x = \frac{\sigma}{\sqrt{n/12}} \sum_{i=1}^{n} r_i + \mu - \frac{n}{2} \frac{\sigma}{\sqrt{n/12}}$$

es una observación aleatoria a partir de una distribución normal aproximada con media μ y desviación estándar σ. Esta aproximación es excelente (excepto en las colas de la distribución), aun para valores pequeños de n. En consecuencia, los valores de n entre 5 y 10 pueden ser adecuados; $n = 12$ también es un valor conveniente porque elimina los términos de la raíz cuadrada de la expresión anterior.

Como siempre se dispone de tablas de la distribución normal (vea el apéndice 5), otro método sencillo para generar una aproximación cercana de una observación aleatoria es usar esa tabla de manera directa para elaborar el método de transformación inversa. Este procedimiento es bastante conveniente cuando se generan varias observaciones a mano, pero lo es menos cuando se programa en computadora puesto que en este caso es necesario almacenar una tabla grande y después usar la tabla de búsqueda.

Se han desarrollado algunas otras técnicas *exactas* para generar observaciones aleatorias a partir de una distribución normal.[4] Estas técnicas exactas son suficientemente rápidas como para que, en la práctica, se usen en lugar del método aproximado que se acaba de describir. Es común que se incorpore una rutina para una de estas técnicas en los paquetes de software con características para simulación. Por ejemplo, Excel usa la función DISTR.NORM.INV(ALEATORIO(), μ, σ), para generar una observación aleatoria que sigue una distribución normal con media μ y desviación estándar σ.

Un método sencillo para manejar la **distribución ji cuadrada** consiste en usar el hecho de que es el resultado de sumar los cuadrados de variables aleatorias normales estandarizadas. Así, si $y_1, y_2, \ldots, y_n$ son n observaciones aleatorias que siguen una distribución normal con media 0 y desviación estándar 1, entonces

$$x = \sum_{i=1}^{n} y_i^2$$

es una observación aleatoria a partir de una distribución ji cuadrada con n grados de libertad.

Método de aceptación-rechazo

En el caso de muchas distribuciones continuas, no es factible aplicar el método de la transformada inversa puesto que quizá $x = F^{-1}(r)$ no se pueda calcular (al menos en forma eficiente). Por lo tanto, se han desarrollado algunos otros tipos de métodos para generar observaciones aleatorias a partir

[4] Vea de nuevo la referencia citada en la nota al pie 3 de la página anterior.

de esas distribuciones. Con frecuencia, estos métodos son mucho más rápidos que el método de la transformada inversa, aun cuando se puede usar este último. Para proporcionar cierta noción del enfoque de estos métodos, se ilustrará uno que recibe el nombre de **método de aceptación-rechazo** con un ejemplo sencillo.

Considere una *distribución triangular* que tiene función de densidad de probabilidad

$$f(x) = \begin{cases} x & \text{si } 0 \leq x \leq 1 \\ 1 - (x - 1) & \text{si } 1 \leq x \leq 2 \\ 0 & \text{de otra manera.} \end{cases}$$

El método de aceptación-rechazo utiliza los siguientes pasos (tal vez varias veces) para generar una observación aleatoria.

1. Generar un número aleatorio uniforme r_i entre 0 y 1 y establecer $x = 2r_i$ (de manera que el intervalo de valores posibles de x es de 0 a 2).
2. Aceptar x con

$$\text{Probabilidad} = \begin{cases} x & \text{si } 0 \leq x \leq 1 \\ 1 - (x - 1) & \text{si } 1 \leq x \leq 2, \end{cases}$$

como la observación aleatoria deseada [puesto que esta probabilidad es igual a $f(x)$]. *De otra manera*, *rechazar* x y repetir los dos pasos.

Para generar en forma aleatoria el evento de aceptar (o rechazar) x de acuerdo con esta probabilidad, el método pone en práctica el paso 2 de la manera siguiente:

3. Generar un número aleatorio uniforme r_2 entre 0 y 1.

Aceptar x si $r_2 \leq f(x)$.
Rechazar x si $r_2 > f(x)$.

Si x se rechaza, es necesario repetir los dos primeros pasos.

Como $x = 2r_1$ se acepta con probabilidad $= f(x)$, la distribución de probabilidad de los *valores aceptados* tiene $f(x)$ como su función de densidad, de manera que los valores aceptados son *observaciones aleatorias* válidas de $f(x)$.

Por fortuna, en este ejemplo el valor *más grande* de $f(x)$ para cualquier x era exactamente 1. Si este valor más grande fuera $L \neq 1$, entonces r_2 quedaría multiplicado por L en el paso 2. Con este ajuste, el método se extiende con facilidad a otras funciones de densidad de probabilidad en un intervalo finito y también se puede usar conceptos similares en un intervalo infinito.

20.5 DESCRIPCIÓN DE UN ESTUDIO DE SIMULACIÓN IMPORTANTE

Hasta ahora, este capítulo se ha centrado en el *proceso* de realizar una simulación y en algunas aplicaciones para hacerlo. Ahora se ubicará este material en una perspectiva más amplia con una descripción breve de los pasos típicos que incluye un estudio importante de investigación de operaciones basado en la aplicación de simulación (se aplican casi los mismos pasos cuando el estudio se basa en otra técnica de investigación de operaciones).

Paso 1: Formulación del problema y planeación del estudio

El equipo de investigación de operaciones debe comenzar por reunirse con la administración para determinar el problema a partir de los siguientes tipos de preguntas.

1. ¿Qué problema desea estudiar la administración?
2. ¿Cuáles son los objetivos globales del estudio?
3. ¿Qué aspectos específicos debe incluirse?
4. ¿Qué tipos de configuraciones de sistemas alternativos deben considerarse?
5. ¿Qué medidas de desempeño del sistema son de interés para la administración?
6. ¿Cuáles son las restricciones de tiempo para realizar el estudio?

Recuadro de aplicación

Los centros de atención telefónica de llamadas han sido una de las industrias de mayor crecimiento a nivel mundial por muchos años. Sólo en Estados Unidos, cientos de miles de negocios usan los centros de atención ubicados en todo el mundo para que los clientes puedan hacer sus pedidos mediante una llamada telefónica sin costo a un número 800.

El mercado de la red 800 es muy lucrativo para las compañías de telecomunicaciones, por lo que están felices de poder vender la tecnología necesaria para sus clientes de negocios y, después, ayudarlos a diseñar centros de atención eficientes. La compañía **AT&T** fue pionera en el desarrollo y comercialización de este servicio para sus clientes. Su enfoque consistió en desarrollar un modelo de simulación altamente sofisticado y flexible llamado *Simulador de Procesamiento de Llamadas (CAPS, Call Processing Simulator)*, que permitía a sus clientes estudiar varios escenarios sobre cómo diseñar y operar sus centros de atención telefónica.

El simulador CAPS consta de cuatro módulos. El *módulo de generación de llamadas* produce llamadas entrantes que llegan de forma aleatoria, con tasas de llegada promedio que varían a lo largo del día. El *módulo de red* simula la forma en que una llamada entrante puede contestarse de manera inmediata o ponerla en espera o recibir una señal de ocupado. En estos últimos casos puede resultar que la persona que llama espere hasta que entre su llamada o que se desespere y haga negocio con otra compañía. El *módulo de distribución automática de llamadas* simula la forma en que el sistema automático de distribución de llamadas de AT&T distribuye de manera equitativa los agentes disponibles. El *módulo de servicio de llamadas* simula que ciertos agentes dan servicio a diferentes llamadas y, después, hacen todo el trabajo de seguimiento necesario.

El desarrollo y perfeccionamiento de CAPS a lo largo de muchos años siguen los pasos de un estudio de simulación muy importante que se describe en la sección 20.5. Este enfoque meticuloso ha dado grandes dividendos a AT&T. La compañía ha realizado alrededor de 2 000 estudios CAPS al año para sus clientes de negocios ayudándole a incrementar, proteger y *ganar más de* **1 000 millones de dólares** en un mercado-de-red-800 de 8 000 millones de dólares. Esto también *ha generado más de* **750 millones de dólares** *de ganancias anuales* para los clientes de negocio de AT&T que solicitaron los estudios acerca de CAPS. Esta sofisticada aplicación de la simulación condujo a AT&T a ganar el primer lugar en la competencia internacional de 1993 por el Premio Franz Edelman al desempeño en investigación de operaciones y ciencias de la administración.

Fuente: A. J. Brigando, D. R. Dargon, M. J. Sheehan y T. Spencer III: "AT&T's Call Processing Simulator (CAPS) Operational Design for Inbound Call Centers", en *Interfaces*, **24**(1): 6-28, enero-febrero de 1994. (En el sitio en internet de este libro —www.mhhe.com/hillier— se proporciona una liga hacia este artículo.)

Además, el equipo deberá reunirse con ingenieros y personal operativo para aprender los detalles de cómo operaría el sistema. (En general, el equipo debe incluir uno o más miembros que tengan un conocimiento directo del sistema.)

Paso 2: Recolección de datos y formulación del modelo de simulación

Los tipos de datos necesarios dependen de la naturaleza del sistema que se simula. En el caso de un sistema de colas, los datos clave serían la distribución de los *tiempos entre llegadas* y de los *tiempos de servicio*. En la mayoría de los demás casos, lo que se necesita son las *distribuciones de probabilidad* de las cantidades relevantes. En general, sólo se podrán estimar estas distribuciones, pero es importante hacerlo. Con la finalidad de generar escenarios representativos de la forma en que un sistema se desempeñará, es fundamental que la simulación genere *observaciones aleatorias* a partir de estas distribuciones en lugar de sólo usar los promedios.

Muchas veces un modelo de simulación se formula en términos de un *diagrama de flujo* que enlaza los componentes del sistema. Las reglas de operación, que se dan para cada componente, incluyen las distribuciones de probabilidad que controlan cuándo ocurrirán los eventos.

Paso 3: Comprobación de la precisión del modelo de simulación

Antes de construir un programa de computadora, el equipo de IO debe comprometer a las personas más familiarizadas con la operación futura del sistema para verificar la precisión del modelo de simulación. Con frecuencia, este compromiso se logra mediante una revisión estructurada del modelo conceptual y su presentación ante una audiencia compuesta por todas las personas clave. En una reunión de este tipo se puede descubrir y corregir algunos supuestos erróneos del modelo, añadir otros y resolver algunos aspectos acerca del grado de detalle necesario de las distintas partes que lo componen.

Paso 4: Selección del software y creación del programa de computadora

Existen varias clases importantes de software que se utilizan para realizar simulaciones en computadora. Uno es la *hoja de cálculo*. El ejemplo 1 de la sección 20.1 ilustra la forma en que Excel puede realizar algunas simulaciones básicas. Además, se dispone de algunos complementos excelentes de Excel para mejorar este tipo de modelado en hoja de cálculo. La siguiente sección se centra en el uso de estos complementos.

Otras clases de software para simulación están diseñadas para aplicaciones más grandes en las que ya no es conveniente usar hojas de cálculo. Una de ellas la constituyen los *lenguajes de programación de propósito general* como C, FORTRAN, BASIC, etc. Estos lenguajes (y sus predecesores) se usaron mucho en los inicios del desarrollo del campo, debido a su gran flexibilidad para programar cualquier tipo de simulación. Sin embargo, por el tiempo tan largo de programación que se requiere, ahora se usan mucho menos.

También se ha desarrollado una gran cantidad de paquetes de software comerciales que no usan hojas de cálculo para llevar a cabo simulaciones específicas. Históricamente, dichos paquetes han sido clasificados en dos categorías: *lenguajes de simulación de propósito general* y *simuladores orientados a la aplicación*. Los primeros proporcionan un gran número de características necesarias para programar cualquier modelo de simulación en una forma eficiente. Los segundos (también llamados *simuladores*) están diseñados para simular muchos tipos específicos de sistemas. Sin embargo, a medida que ha pasado el tiempo, la diferencia entre estas dos categorías casi ha desaparecido. Los lenguajes de programación de propósito general incluyen ciertas características especiales que los hacen apropiados para trabajar como simuladores en ciertos tipos de aplicaciones específicas. Por el contrario, los simuladores actuales tienden a incluir un mayor grado de flexibilidad en el sentido de su capacidad para trabajar con una clase de sistemas más amplia.

Otra forma de categorizar a los paquetes de simulación es en función al uso de un enfoque de *programación por evento* o un *enfoque por proceso para la modelación* de la simulación de eventos discretos. El primer enfoque va muy de la mano con el método del a*vance de tiempo incremental del próximo evento* que se describió en la sección 20.1. El *enfoque por proceso* también usa como fondo el incremento del próximo evento, pero se enfoca en el modelado en lugar de hacerlo en la descripción de los procesos que generan los eventos. La mayoría de los paquetes de simulación actuales usan el enfoque por proceso.

Cada vez es más común que los paquetes de simulación incluyan características de **animación** para desplegar simulaciones en acción. En una animación, los elementos clave de un sistema se representan en una pantalla de computadora mediante íconos que cambian su forma, color o posición cuando se presenta un cambio en el estado del sistema de simulación. La razón primordial de la gran popularidad de la animación es su capacidad para comunicar la esencia de un modelo de simulación (o de una corrida de simulación) a los administradores y otras personas clave.

Debido a la creciente importancia de la simulación, existen en la actualidad alrededor de 50 compañías de software que comercializan paquetes de simulación. La referencia seleccionada 12 proporciona un estudio de dichos paquetes. (*OR/MS Today* actualiza dicho estudio cada dos años.)

Paso 5: Prueba de validación del modelo de simulación

Después de construir y depurar el programa de computadora, el siguiente paso es probar si el modelo de simulación incorporado en él proporciona resultados válidos para el sistema que intenta representar. En particular, ¿los valores de las medidas generadas por el modelo de simulación son una aproximación cercana de las medidas de desempeño del sistema real?

En algunos casos suele disponerse de un modelo matemático que proporcione resultados de una versión sencilla del sistema. Si es así, estos resultados deben compararse con los resultados de la simulación.

Cuando no se dispone de datos reales para comparar con los resultados de la simulación, una posibilidad es realizar una *prueba de campo* para recolectarlos. Esta prueba incluye construir un pequeño prototipo de alguna versión del sistema propuesto y ponerlo en marcha.

Otra prueba de validación útil es que el personal operativo experimentado verifique la credibilidad del cambio en los resultados de la simulación cuando cambia la configuración del sistema simulado. La observación de animaciones de corridas de simulación también constituye una forma muy útil de verificar la validez del modelo de simulación.

Paso 6: Planeación de las simulaciones que deben realizarse

En este punto, es necesario comenzar a tomar decisiones acerca de las configuraciones del sistema que se va a simular. A menudo, éste es un proceso evolutivo, donde los resultados iniciales de una gama de configuraciones ayudan a determinar qué configuraciones específicas justifican una investigación detallada.

También deben tomarse decisiones sobre algunos aspectos estadísticos. Uno de ellos (a menos que se use la técnica especial que se describió en el complemento a este capítulo en el sitio en internet de este libro) es la *longitud del periodo de calentamiento* para esperar que el sistema alcance, en esencia, la condición de estado estable antes de iniciar la recolección de datos. Con frecuencia, las corridas preliminares se usan para analizar este aspecto. Como los sistemas suelen requerir un tiempo sorprendentemente largo para llegar al estado estable, es útil seleccionar *condiciones iniciales* de un sistema simulado que parezca representativo de las condiciones de estado estable a fin de reducir al mínimo este tiempo requerido.

Otro aspecto estadístico importante es la *longitud de la corrida de simulación* después del periodo de calentamiento de cada configuración del sistema que se simula. Es necesario tener presente que la simulación no produce valores exactos de las medidas de desempeño. En su lugar, cada corrida de simulación puede verse como un *experimento estadístico* que genera *observaciones estadísticas* del desempeño del sistema simulado. Estas observaciones se usan para obtener *estimaciones estadísticas* de las medidas de desempeño. Cuando aumenta la longitud de una corrida se incrementa la precisión de estas estimaciones. (También, en el primer complemento de este capítulo en el sitio en internet de este libro se describen *técnicas especiales de reducción de la varianza* que algunas veces pueden usarse para incrementar la precisión de estas estimaciones.)

La teoría estadística para diseñar experimentos estadísticos que se realizan mediante simulación es muy similar a la de los experimentos realizados por observación directa del desempeño de un sistema físico.[5] Por lo tanto, la inclusión de un estadístico profesional (o al menos un analista con experiencia en simulación y antecedentes importantes en estadística) en el equipo de IO puede ser invaluable en este paso.

Paso 7: Realización de corridas de simulación y análisis de los resultados

La salida de la corrida de simulación proporciona estimaciones estadísticas de las medidas de desempeño deseadas de cada configuración del sistema que interesa. En general, además de una *estimación puntual* de cada medida de desempeño debe obtenerse un *intervalo de confianza* para indicar los valores probables de la medida (como se hizo en el ejemplo 2 de la sección 20.1). En el segundo complemento de este capítulo en el sitio en internet del libro se describe un método para hacerlo.[6]

Estos resultados pueden indicar de inmediato que es claro que una configuración del sistema es superior a las otras. Con mayor frecuencia identificarán unos cuantos candidatos fuertes para ser el mejor. En el último caso, deben realizarse algunas corridas de simulación más largas para comparar mejor estos candidatos.[7] También pueden usarse corridas adicionales para afinar los detalles de la que parece la mejor configuración.

Paso 8: Presentación de recomendaciones a la administración

Después de completar su análisis, el equipo de IO debe presentar sus recomendaciones a la administración. Esta presentación suele hacerse mediante un informe por escrito y una presentación formal a los administradores responsables de tomar las decisiones respecto del sistema que se estudia.

---

[5] Para consultar detalles acerca de teoría estadística relevante para la aplicación de la simulación, vea los capítulos 9 a 12 en la referencia seleccionada 10. Asimismo, consulte las referencias seleccionadas 8 y 9 para ver estudios de gran importancia acerca del diseño y análisis de experimentos de simulación.

[6] Consulte las páginas 530 y 531 de la referencia seleccionada 10 para ver métodos alternos.

[7] La metodología para utilizar simulación para intentar identificar la mejor configuración del sistema se conoce como *optimización de la simulación*. En la actualidad, ésta es un área muy activa de investigación. Por ejemplo, consulte las referencias seleccionadas 7, 13 y 4.

El informe y la presentación, además de resumir la manera en que se realizó el estudio, deben incluir documentación que valide el modelo de simulación. También se puede incluir una *animación* de una corrida simulada para comunicar mejor el proceso de simulación y agregar credibilidad. Asimismo, es importante mostrar los resultados numéricos que proporcionan la base lógica de las recomendaciones.

Por lo general, la administración involucra al equipo de IO en la implantación inicial del nuevo sistema, lo cual incluye la capacitación del personal afectado.

20.6 SIMULACIÓN CON HOJAS DE CÁLCULO

En la sección 20.5 se describen los elementos clásicos de un estudio de simulación de sistemas complejos, entre los cuales se incluye el uso de los lenguajes generales de simulación o simuladores especiales para estudiar la mayoría de estos sistemas de manera eficiente. Sin embargo, no todos los estudios de simulación son tan elaborados. En realidad, cuando se estudian sistemas relativamente sencillos, en ocasiones es posible correr una simulación rápida y sencilla en una hoja de cálculo. En particular, siempre que se usa un modelo en hoja de cálculo para analizar un problema sin tomar en cuenta las incertidumbres (excepto mediante el análisis de sensibilidad), es posible extender el modelo para usar simulación y considerar el efecto de ellas. Por lo tanto, se hará hincapié en estos casos más simples en los cuales se pueden usar hojas de cálculo para realizar simulaciones en forma eficaz.

Como se ilustra en el ejemplo 1 de la sección 20.1, el paquete estándar de Excel tiene algunas características básicas de simulación que incluyen la posibilidad de generar números aleatorios uniformes y observaciones aleatorias a partir de una distribución de probabilidad. Un progreso interesante en años recientes ha sido el desarrollo de algunos complementos de simulación para Excel que extienden en gran medida estas capacidades. Uno de estos complementos es *Crystal Ball*, desarrollado por Decisioneering, Inc. (Oracle en la actualidad). Además de su gran funcionalidad para llevar a cabo simulaciones, la edición profesional de Crystal Ball también incluye dos módulos adicionales. Uno es el CB Predictor, que sirve para generar pronósticos a partir de datos de las series de tiempo, como se describe e ilustra en el capítulo 27 (un capítulo complementario en el sitio en internet de este libro). El otro es OptQuest, el cual supera a CrystalBall pues utiliza el resultado de una serie de corridas de simulación para buscar en forma automática una solución óptima para un modelo de simulación, como se describe en el tercer complemento de este capítulo en el sitio en internet de este libro.

También se dispone de otros complementos de simulación como software liberado. Uno de ellos es RiskSim, desarrollado por el profesor Michael Middleton. La versión académica de RiskSim se incluye en el OR Courseware. Aunque no es tan elaborado o poderoso como Crystal Ball, RiskSim es fácil de usar y está bien documentado en el sitio en internet de este libro. (Si desea continuar usándolo más adelante, debe registrarse y pagar la cuota de liberación.) Como cualquier complemento de Excel, cualquiera de estos programas de simulación debe ser instalado antes de que Excel los pueda mostrar.

Esta sección se dedica al uso de Crystal Ball para ilustrar lo que se puede hacer con los complementos de simulación. Usted puede obtener práctica en el uso de Crystal Ball consultando su sitio en internet (www.decisioneering.com/downloadform.html) para bajar este software por un periodo de prueba (30 días en la actualidad). Su universidad (como muchas otras) puede contar con una licencia de este popular paquete de software.

Al final del capítulo se han incluido problemas para esta sección y la próxima, con los que es posible utilizar Crystal Ball. También se puede usar RiskSim, que se incluye en el sitio en internet de este libro, para resolver los problemas de estas secciones.

Las hojas de cálculo de negocios casi siempre incluyen algunas *celdas de entrada* que despliegan datos clave (como los costos asociados con la producción o comercialización de un producto) y una o más *celdas de salida* que muestran las medidas de desempeño (como la ganancia de producir o vender el producto). El usuario escribe ecuaciones de Excel para enlazar las entradas con las salidas de manera que éstas muestren los valores correspondientes a los que se introdujo. En algunos casos, habrá incertidumbre respecto de cuáles serán los valores de entrada correctos. Se puede utilizar el análisis de sensibilidad para verificar el cambio en la salida cuando cambian las celdas de entrada. Si existe una incertidumbre considerable en cuanto a los valores de algunas celdas de

entrada, será útil adoptar un enfoque más sistemático para analizar el efecto de esa incertidumbre. Aquí es donde la simulación aparece en el horizonte.

Cuando se utiliza la simulación, en lugar de introducir un solo número en una celda de entrada cuando hay incertidumbre, se proporciona una *distribución de probabilidad* que la describe. Al generar una *observación aleatoria* que sigue una distribución para cada celda de entrada de este tipo, la hoja calcula los valores de salida en la forma usual. En Crystal Ball, este proceso se llama **iteración**. Al correr el número de iteraciones especificadas por el usuario (cientos o miles), la simulación genera ese número de observaciones aleatorias de los valores de salida. El programa Crystal Ball registra toda esta información y después le da la opción de imprimir una tabla o una gráfica (o ambas) con las estadísticas detalladas que muestra la *distribución de probabilidad* de los valores de salida. Un resumen de los resultados incluye también estimaciones de la media y la desviación estándar de esta distribución.

Ahora se verá un ejemplo en detalle para ilustrar este proceso.

Ejemplo de administración de inventarios: el problema de Freddie, el voceador

Considere el siguiente problema que enfrenta Freddie, el voceador. Uno de los periódicos que vende en su puesto es el *Financial Journal*. Cada mañana, un distribuidor entrega al puesto los ejemplares diarios del periódico. Cada ejemplar que no se vende durante el día se devuelve al distribuidor la siguiente mañana. Sin embargo, para estimular la compra de un número grande de ejemplares, el distribuidor da a Freddie un pequeño reembolso por los ejemplares no vendidos.

A continuación se presentan los costos de Freddie.

Freddie paga $1.50 por ejemplar entregado.
Freddie lo vende en $2.50.
Freddie recibe un reembolso de $0.50 por cada uno que no vende.

En parte por el reembolso, Freddie siempre ha pedido un suministro grande. Sin embargo, está preocupado pues paga mucho por los ejemplares que tiene que devolver por no haber sido vendidos, en particular porque esto ha sucedido casi a diario. Él piensa que sería mejor ordenar sólo un número mínimo de ejemplares y ahorrarse este costo extra.

Para investigar esta situación más a fondo, ha recopilado el siguiente registro de sus ventas diarias.

Freddie vende cualquier cantidad entre 40 y 70 ejemplares en cualquier día de la semana.
La frecuencia de las cantidades entre 40 y 70 es aproximadamente la misma.

La decisión que Freddie quiere tomar es determinar el número de ejemplares que debe ordenar al distribuidor por día. Su objetivo es maximizar la ganancia diaria promedio.

Este problema se puede reconocer como un ejemplo del *problema del voceador* que se presentó en la sección 18.7. En consecuencia, se puede usar el *modelo estocástico de inventario de un solo producto perecedero* (sin costo de preparación) para resolver este problema. Sin embargo, con propósitos ilustrativos, se mostrará cómo se puede usar la simulación para analizar este sencillo sistema de inventarios, de la misma manera en que se analizan sistemas de inventarios más complejos que están fuera del alcance de los modelos disponibles.

Modelo en hoja de cálculo para este problema

En la figura 20.7 se muestra un modelo en hoja de cálculo para este problema. Dadas las celdas de datos C4:C6, la variable de decisión es la cantidad a ordenar que se introduce en la celda C9. (Se ha escogido la cantidad de 60 como una primera estimación de un valor razonable.) En la parte baja de la figura se muestran las ecuaciones que se usan para calcular las celdas de salida C14:C16. Estas celdas de salida son usadas para calcular la celda de salida Ganancia (C18).

La única cantidad de entrada incierta en esta hoja de cálculo es la demanda del día en la celda C12. Esta cantidad puede estar en cualquier punto entre 40 y 70. Debido a que las frecuencias de los números entre 40 y 70 son casi las mismas, la distribución de probabilidad de la demanda diaria puede suponerse en forma razonable como una *distribución uniforme* entre 40 y 70, como se indica en las celdas D12:F12. En lugar de introducir un solo número de manera permanente en Deman-

	A	B	C	D	E	F
1	**Freddie, el voceador**					
2						
3			**Datos**			
4		Precio de venta unitario	$ 2.50			
5		Costo de compra unitario	$1.50			
6		Valor de salvamento unitario	$0.50			
7						
8			**Variable de decisión**			
9		Cantidad por ordenar	60			
10						
11			**Simulación**		Mínimo	Máximo
12		Demanda simulada	55	Uniforme	40	70
13		Demanda (redondeada)	55			
14						
15		Ingreso por ventas	$ 2.50			
16		Costo de compra	$1.50			
17		Valor de salvamento	$0.50			
18						
19		Ganancia	$50.00			

	B	C
11		**Simulación**
12	Demanda simulada	55
13	Demanda (redondeada)	=REDONDEAR(DemandaSimulada, 0)
14		
15	Ingreso por ventas	=PrecioDeVentaUnitario*MIN(CantidadAOrdenar,Demanda)
16	Costo de compra	=CostoDeCompraUnitario*CantidadAOrdenar
17	Valor de salvamento	=ValorDeSalvamentoUnitario*MAX(CantidadAOrdenar-Demanda,0)
18		
19	Ganancia	=IngresoPorVentas-CostodeCompra+ValordeSalvamento

Nombre de rango	Celda
Demanda	C13
CantidadAOrdenar	C9
Ganancia	C19
CostoDeCompra	C16
IngresoPorVentas	C15
IngresoPorVentas	C17
DemandaSimulada	C12
CostoDeCompraUnitario	C5
PrecioDeVentaUnitario	C4
ValorDeSalvamentoUnitario	C6

■ **FIGURA 20.7**
Modelo en hoja de cálculo para aplicar simulación al ejemplo de Freddie, el voceador. La celda de supuesto es DemandaSimulada (C12), la celda de pronóstico es Ganancia (C18), y la variable de decisión es CantidadA Ordenar (C9).

daSimulada (C12), Crystal Ball introducirá esta distribución de probabilidad en la celda. (Antes de entrar en Crystal Ball, se introduce en esta celda un valor temporal de 55, como se muestra en la figura 20.7.) En razón de que con el uso de Crystal Ball se genera una *observación aleatoria* a partir de esta distribución de probabilidad, la hoja de cálculo puede calcular la celda de salida en el modo usual para completar una iteración. Si se corre el número de iteraciones especificada por el usuario (de manera normal cientos o miles), la simulación genera el mismo número de observaciones aleatorias de los valores en las celdas de salida. Crystal Ball registra esta información de la(s) celda(s) de salida de interés particular (ganancia diaria de Freddie) y después, al final, la despliega en una variedad de formas convenientes que revelan una estimación de la distribución de probabilidad subyacente de la ganancia diaria de Freddie. (Más adelante, en este capítulo, se examinará más de cerca este aspecto.)

Aplicación de Crystal Ball

Se deben dar cuatro pasos para usar la hoja de cálculo de la figura 20.7 para ejecutar la simulación con Crystal Ball.

1. Definir las celdas de entrada aleatorias.
2. Definir las celdas de salida a pronosticar.

3. Establecer las preferencias de la corrida.
4. Correr la simulación.

A continuación se describe cada uno de estos pasos.

Definición de las celdas de entrada aleatorias. Una celda de entrada aleatoria es una celda de entrada que tiene un valor aleatorio (como la demanda diaria del *Financial Journal*), por lo cual se requiere introducir en la celda una distribución de probabilidad supuesta en lugar de introducir en forma permanente un solo número. La única celda de entrada aleatoria de la figura 20.7 es DemandaSimulada (C12). Crystal Ball hace referencia a cada una de estas celdas de entrada aleatoria como **celda de supuesto** (*assumption cell*).

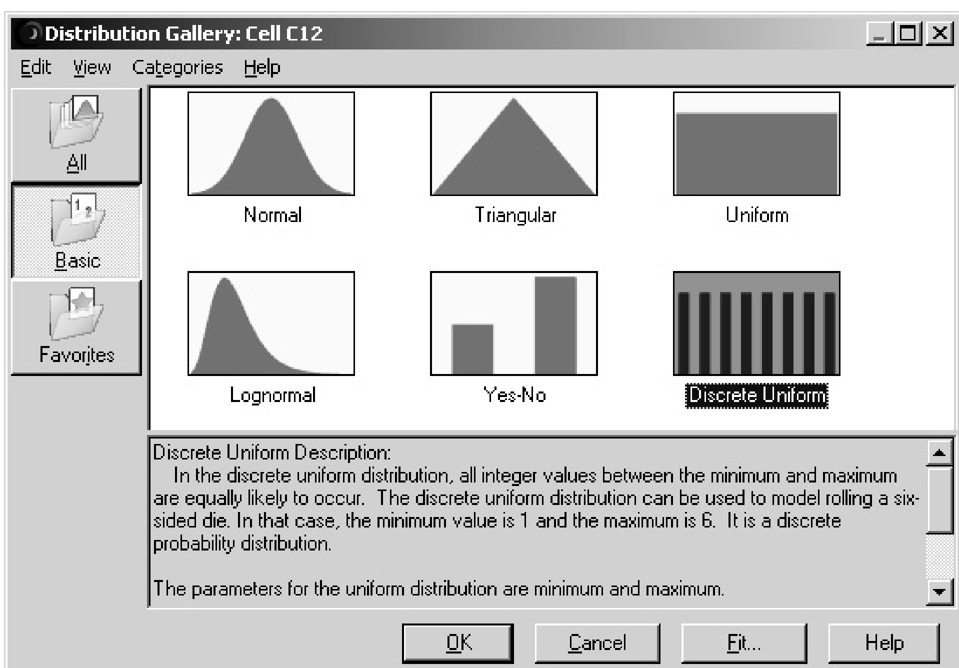
El siguiente procedimiento se utiliza para definir una celda de supuesto.

Procedimiento para definir una celda de supuesto

1. Seleccione la celda al hacer clic sobre ella.
2. Si la celda no contiene un número, introduzca *cualquier* valor en ella.
3. Haga clic sobre el botón Define Assumption  en la pestaña Crystal Ball (de Excel 2007) o barra de herramientas (en el caso de versiones anteriores de Excel).
4. Seleccione una distribución de probabilidad para introducir en la celda al hacer clic sobre la distribución escogida de la galería de distribuciones que se muestra en la figura 20.8.
5. Haga clic en OK (o doble clic sobre la distribución) para llamar el cuadro de diálogo de la distribución seleccionada.
6. Use este cuadro de diálogo para introducir los parámetros de la distribución; preferentemente, haga referencia a las celdas de la hoja de cálculo que contienen los valores de estos parámetros. Si lo desea, también puede introducir un nombre para la celda de supuesto. (Si la celda ya tiene un nombre, el mismo aparecerá en el cuadro de diálogo.)
7. Haga clic en OK.

La **galería de distribuciones** que se mencionó en el paso 4 proporciona 21 diferentes distribuciones de probabilidad entre las cuales se puede elegir. En la figura 20.8 se despliegan seis

■ **FIGURA 20.8**
El cuadro de diálogo de Distribution Gallery de Crystal Ball muestra las distribuciones básicas. Además de las 6 distribuciones que se presentan aquí, puede accederse a 15 más haciendo clic en el botón "All".



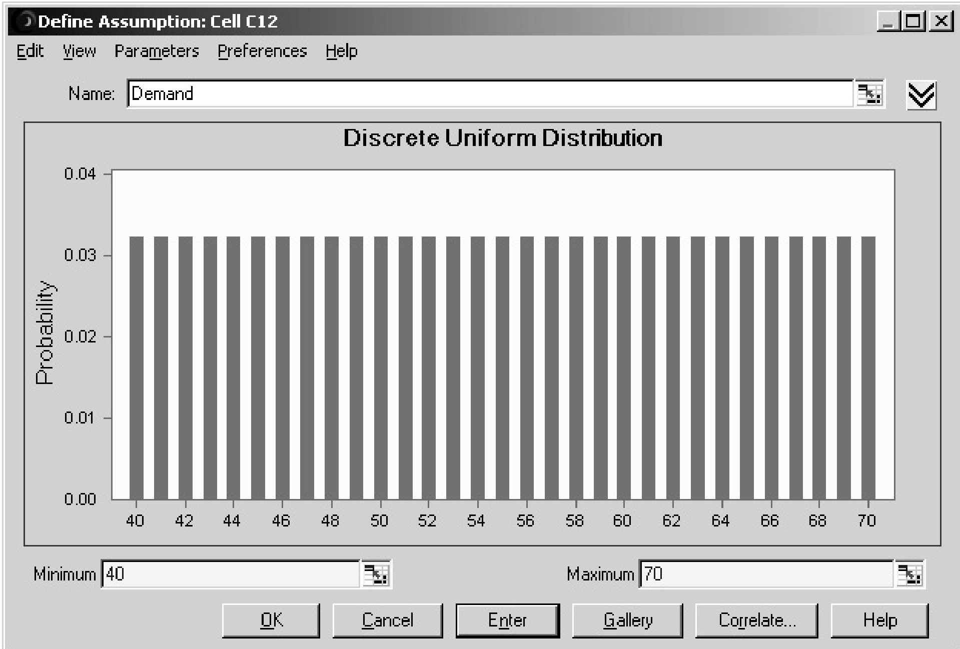


FIGURA 20.9
Cuadro de diálogo de la distribución uniforme discreta de Crystal Ball. En este caso se utiliza para ingresar una distribución uniforme discreta con los parámetros 40(=E12) y 70(=F12) en la supuesta celda Demand (C12) en el modelo de hoja de cálculo de la figura 20.7.

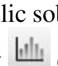
distribuciones básicas, pero hay 15 más disponibles al hacer clic sobre el botón “All”. (Cuando existe incertidumbre acerca de cuál distribución continua proporciona el mejor ajuste a los datos históricos, Crystal Ball cuenta con una rutina para escoger una distribución apropiada. Esta rutina se describe en la sección 28.6 del sitio en internet de este libro.)

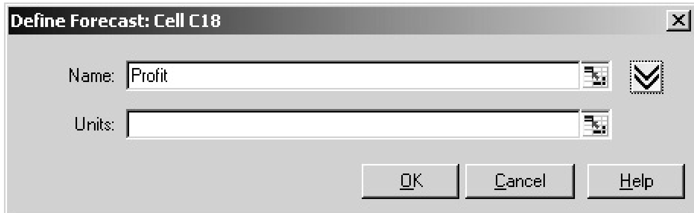
En el caso de Freddie, al hacer doble clic sobre la distribución uniforme de la galería de distribuciones, surge el cuadro de diálogo de la distribución uniforme que se muestra en la figura 20.9. Este cuadro se utiliza para introducir los parámetros de la distribución. Para cada uno de estos parámetros (Minimum y Maximum) se hace referencia a las celdas en E12 y F12 de la hoja de cálculo escribiendo las fórmulas =E12 y =F12 para Minimum y Maximum, respectivamente. Después de ingresar las referencias de las celdas, aparecerá un cuadro de diálogo que muestra el valor real del parámetro con base en la referencia de las celdas (40 y 70 como se muestra en la figura 20.9). Para ver o realizar un cambio a una referencia de celdas, presione en el parámetro a fin de que se despliegue la referencia a las celdas correspondiente.

Definición de las celdas de salida a pronosticar. Crystal Ball se refiere a la salida de una simulación como un *pronóstico*, puesto que pronostica cuál será la función de probabilidad del desempeño del sistema real cuando éste comience a operar. Por lo tanto, cada celda de salida que se utiliza en una simulación para pronosticar una medida de desempeño se llama **celda de pronóstico**. El modelo de hoja de cálculo para una simulación en computadora no incluye una celda objetivo, pero las celdas de pronóstico tienen un papel parecido.

La medida de desempeño que interesa a Freddie es su ganancia diaria por vender el periódico, por lo que la única celda de pronóstico de la figura 20.7 es Ganancia (C18). A continuación se presenta el procedimiento para definir una celda de salida como una de pronóstico.

Procedimiento para definir una celda de pronóstico

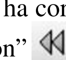
1. Seleccione la celda por medio de un clic sobre ella.
2. Haga clic en el botón Define Forecast  en la pestaña de Crystal Ball (en el caso de Excel 2007) o en la barra de herramientas (en el caso de versiones anteriores de Excel), con lo que se activa el cuadro de diálogo Define Forecast (como se muestra en la figura 20.10 del problema de Freddie).



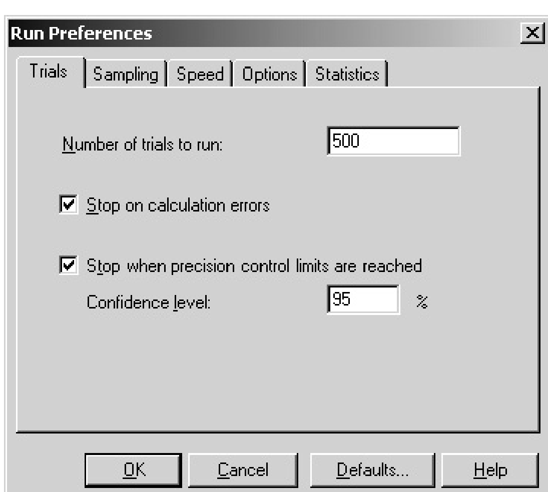
■ **FIGURA 20.10**
Cuadro de diálogo para definir pronóstico (Define Forecast) en Crystal Ball. En este caso se utiliza para definir la celda de pronóstico Profit (C18) en el modelo de hoja de cálculo de la figura 20.7.

3. Este cuadro de diálogo se puede utilizar para definir un nombre y (de manera opcional) unidades para la celda de pronóstico. (Si ya se ha asignado un nombre a la celda, el mismo aparecerá en el cuadro de diálogo.)
4. Haga clic en OK.

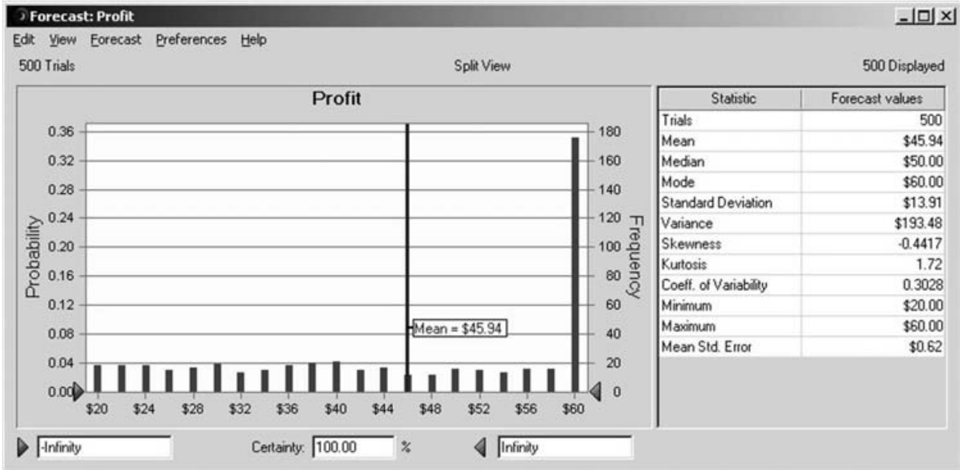
Establecimiento de preferencias para la corrida. El tercer paso —establecer las preferencias de la corrida— se refiere a acciones como elegir el número de iteraciones y decidir sobre otras opciones en relación a la forma de realizar la simulación por computadora. Este paso se inicia al hacer clic sobre el botón "Run Preferences" en la pestaña de Crystal Ball (Excel 2007) o en la barra de herramientas (en el caso de versiones anteriores de Excel). El cuadro de diálogo de preferencias de la corrida tiene las cinco opciones que se muestran en la parte superior de la figura 20.11. Se puede hacer clic en cualquiera de estos botones para introducir algún cambio a las especificaciones controladas por dicha opción acerca de cómo correr la simulación. Por ejemplo, en la figura 20.11 se muestra la versión del cuadro de diálogo que se obtiene al seleccionar el botón correspondiente a las iteraciones ("Trial"). Esta figura indica que se ha elegido 500 como el número máximo de iteraciones para la simulación en computadora. (La segunda opción en este cuadro de diálogo, en la que se especifica detenerse si se ha alcanzado la precisión especificada, se describirá después.)

Corrida de simulación. En este punto todo está listo para comenzar a correr la simulación. Para iniciar, sólo es necesario hacer clic en el botón "Start Simulation"  (vea la parte media de la figura 20.8) o elegir "Run Simulation" del menú "Run". Sin embargo, si se ha corrido una simulación previamente, primero se debe hacer clic sobre el botón "Reset Simulation"  o elegir "Reset Simulation" del menú "Run" para borrar la simulación antes de comenzar con la nueva.

Una vez iniciada, se despliega una ventana de pronóstico en la que se muestran los resultados de la simulación al mismo tiempo que ésta corre. En la figura 20.12 se presenta el pronóstico de



■ **FIGURA 20.11**
Cuadro de diálogo Run Preferences de Crystal Ball después de seleccionar la pestaña Trials (iteraciones).



■ **FIGURA 20.12**
Gráfica de frecuencia y tabla de estadísticas proporcionadas por Crystal Ball para resumir los resultados de la corrida del modelo de simulación en la figura 20.7 en el ejemplo de Freddie, el voceador.

GananciaTotal (la ganancia diaria de Freddie por vender el *Financial Journal*) después de que se han completado las 500 iteraciones. La vista predeterminada del pronóstico es la gráfica de frecuencia que se muestra en la mitad superior de la figura. La altura de las líneas verticales de la gráfica de frecuencia indica la frecuencia relativa de los diferentes valores de ganancia que se obtuvieron durante la corrida de simulación. Por ejemplo, considere la línea vertical alta en 60 dólares. En ese punto, el lado derecho de la gráfica indica una frecuencia de 175, lo que significa que 175 de las 500 iteraciones condujeron a una ganancia de 60 dólares. Por lo tanto, el lado izquierdo de la gráfica indica que la probabilidad estimada de una ganancia de 60 dólares es $175/500 = 0.35$. Ésta es la ganancia que resulta siempre que la demanda es igual o superior a la cantidad ordenada de 60. El resto del tiempo, la ganancia se distribuye en forma bastante uniforme entre 20 y 60 dólares. Estos valores de ganancia corresponden a iteraciones donde la demanda está entre 40 y 60 unidades, con valores de ganancia más bajos correspondientes a demandas más cercanas a 40 y valores de ganancia más altos para demandas más cercanas a 60. La media de los 500 valores de ganancia es de \$45.94, como lo indica la *línea media* (la línea vertical punteada) en este punto.

La mitad inferior de la figura 20.12 muestra la tabla que se obtiene al elegir las medidas estadísticas ("Statistics") del menú "View". Estas estadísticas resumen el resultado de las 500 iteraciones de la simulación, las cuales proporcionan una muestra de 500 observaciones aleatorias a partir de la distribución de probabilidad subyacente de la ganancia diaria de Freddie. Las estadísticas más interesantes acerca de esta muestra proporcionadas por la tabla incluyen la *media* de \$45.94, la *mediana* de \$50.00 (que indica el punto medio de las 500 iteraciones cuando se ordenan las ganancias desde la menor hasta la mayor), la *moda* de 60 dólares (que significa que éste fue el valor de ganancia que ocurrió con mayor frecuencia), y la *desviación estándar* de \$13.91. La información al final de la tabla, en relación con el *rango* de valores de ganancia, también resulta particularmente útil.

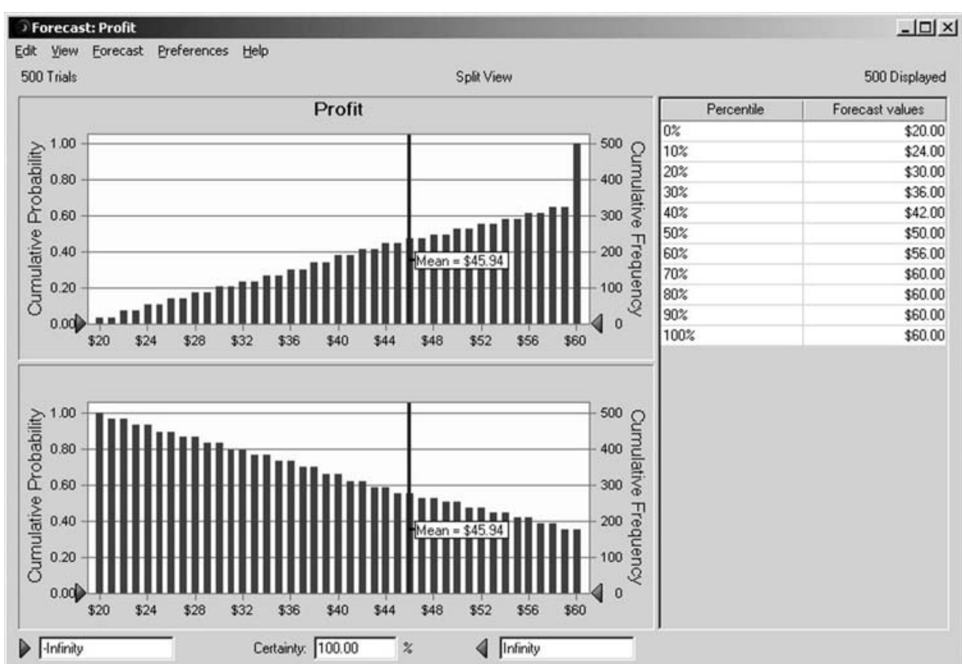
En realidad, la importancia particular de cada una de las estadísticas de la figura 20.12 depende de lo que quiera lograr Freddie. Por lo general, la media es la más importante puesto que, a pesar de las amplias fluctuaciones en las ganancias diarias, la ganancia diaria promedio convergirá hacia la media a lo largo del tiempo. Por lo tanto, si se multiplica la media por el número de días que estará abierto el puesto de periódicos durante el año, se obtiene (de manera muy cercana) cuál será la ganancia anual por vender el periódico, que es una cantidad muy relevante que se desea maximizar. Sin embargo, si Freddie es un individuo que se enfoca más en el presente que en el futuro, entonces la mediana y la moda pueden tener un interés considerable para él. Si él considera que con una ganancia de 50 dólares ha tenido un buen día, y su meta es tener un día al menos como éste la mitad de las veces, entonces él querría que la mediana fuera de al menos 50 dólares (como lo es). Si él obtiene una satisfacción particular por alcanzar la ganancia posible máxima de 60 dólares (dada una cantidad a ordenar de 60 dólares), entonces querrá estar seguro que esto ocurrirá con mayor frecuencia que cualquier otra ganancia específica (como lo indica la moda de 60 dólares). Por otro lado, si Freddie tiene aversión al riesgo y por ende está particularmente preocupado por

evitar los días malos (ganancias muy por debajo de la media) tanto como sea posible, entonces tendría un interés especial en obtener una desviación estándar relativamente baja y un mínimo relativamente grande.

Tenga en mente que las estadísticas de la figura 20.12 se basan en el uso de una cantidad ordenada de 60, mientras que el objetivo es determinar la mejor cantidad a ordenar. Si Freddie tiene un interés particularmente fuerte en más de una de las estadísticas, una posibilidad sería regresar al modelo de simulación de la figura 20.12 con diferentes cantidades por ordenar y dejar que Freddie elija aquella que coloque las estadísticas de interés en el valor que más le guste. Sin embargo, en la mayoría de las situaciones la media será la estadística con un interés especial. En este caso, el objetivo es determinar la cantidad por ordenar que maximiza la media. (De aquí en adelante se supondrá este objetivo.) Después de estimar la cantidad óptima a ordenar de acuerdo con este objetivo, se le debe mostrar a Freddie la gráfica de frecuencia correspondiente y la tabla de estadísticas (y quizá otra información que se describirá después) para estar seguro de que todo lo demás es satisfactorio con esta cantidad a ordenar.

Además de la gráfica de frecuencia y la tabla de estadísticas que se presentaron en la figura 20.12, el menú "View" proporciona algunas otras formas útiles de desplegar los resultados de una corrida de simulación, que incluyen una tabla de percentiles, una gráfica acumulada y una gráfica acumulada inversa. Estos despliegues alternativos se muestran en la figura 20.13. La tabla de percentiles se basa en la ordenación de los valores de ganancia que generaron las 500 iteraciones desde el menor hasta el mayor, después esta lista se divide en 10 partes iguales (50 valores en cada una), y por último se registra el valor al final de cada parte. En consecuencia, el valor en 10% de la lista es 24 dólares, el valor en 20% de la lista es 30 dólares, y así sucesivamente. (Por ejemplo, la interpretación intuitiva del percentil 10 de 24 dólares es que existe 10% de las iteraciones con valores de ganancia menores o iguales a 24 dólares y que el otro 90% de las iteraciones tiene valores de ganancia mayores o iguales, así que 24 dólares es la línea de división entre 10% de los valores más pequeños y 90% de los valores más grandes.) La gráfica acumulada en la parte superior izquierda de la figura 20.13 proporciona información similar (pero más detallada) acerca de la misma lista de valores de ganancia que va desde los valores más bajos a los más altos. El eje horizontal muestra el rango completo de valores desde el valor de ganancia más pequeño posible (20 dólares) hasta el más grande (60 dólares). Para cada valor en este rango, la gráfica acumula el número de ganancias reales generadas por las 500 iteraciones que son menores o iguales a dicho valor. El número es igual a la frecuencia que se muestra a la derecha o, si se divide entre el número

■ **FIGURA 20.13**
Tres formas adicionales en las que Crystal Ball despliega los resultados de la corrida del modelo de simulación en la figura 20.7 en el ejemplo de Freddie, el voceador.



de iteraciones, la probabilidad que se muestra a la izquierda. La gráfica acumulada inversa en la parte inferior izquierda de la figura 20.13 se construye de la misma forma que la gráfica acumulada excepto por la siguiente diferencia crucial. Para cada valor en el rango de 20 y 60 dólares, la gráfica acumulada inversa acumula el número de ganancias reales generadas por las 500 iteraciones que son *mayores* o iguales a dicho valor.

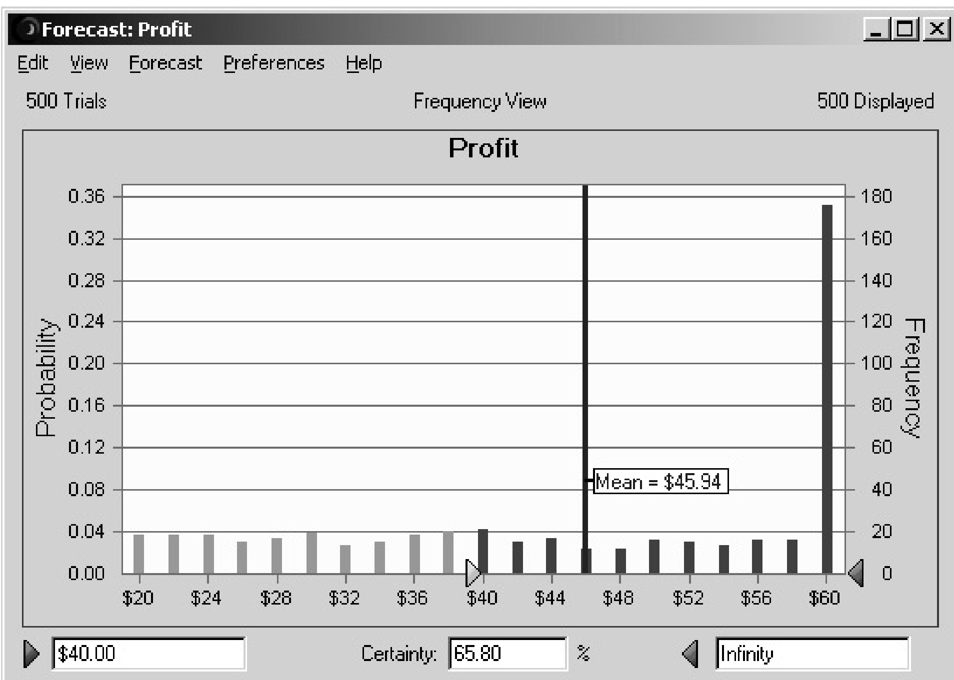
En la figura 20.14 se presenta otra de las muchas formas proporcionadas por Crystal Ball para extraer información útil a partir de los resultados de una corrida de simulación. Freddie siente que ha tenido un día razonablemente satisfactorio si obtiene una ganancia de al menos 40 dólares por ventas del *Financial Journal*. Por lo tanto, le gustaría saber el porcentaje de días que podría esperar que se lograra dicha ganancia si adopta la cantidad por ordenar que se analiza en este momento (60). En el cuadro de certidumbre que está debajo de la gráfica de frecuencia de la figura 20.14, se muestra una estimación de este porcentaje (65.80%). Crystal Ball puede proporcionar este porcentaje de dos maneras. Primero, el usuario puede arrastrar el triángulo que se encuentra a la izquierda justo debajo de la gráfica (originalmente en 20 dólares en la figura 20.12), hacia la derecha hasta llegar a 40 dólares (como en la figura 20.14). De manera alterna, se puede introducir 40 dólares directamente en el cuadro, en la esquina inferior izquierda. Si se desea, también se puede estimar de inmediato la probabilidad de obtener una ganancia entre cualesquiera dos valores al arrastrar los dos triángulos hacia dichos valores.

¿Cuán exactos son los resultados de la simulación?

Un número importante que proporciona la figura 20.12 es la media de $45.94. Este número se calculó como el *promedio* de las 500 observaciones aleatorias a partir de la distribución de probabilidad subyacente de la ganancia diaria de Freddie, las cuales fueron generadas por las 500 iteraciones. Este *promedio de la muestra* proporciona una *estimación* de la *media verdadera* de esta distribución. Sin embargo, la media verdadera se puede desviar en alguna medida de $45.94. ¿Cuán exacta se puede esperar que sea esta estimación?

La respuesta a esta pregunta clave la proporciona el *error estándar de la media* de $0.62 que se proporciona al final de la tabla de estadísticas de la figura 20.12. El error estándar de la media se calcula como $s/\sqrt{n}$, donde s es la desviación estándar de la muestra y n es el número de iteracio-

■ **FIGURA 20.14**
Después de establecer una cota inferior de 40 dólares para los valores de ganancia deseables, el cuadro de certidumbre ("Certainty") debajo de esta gráfica de frecuencia revela que 65.80% de las iteraciones en la corrida de simulación de Freddie proporcionó al menos esta ganancia.

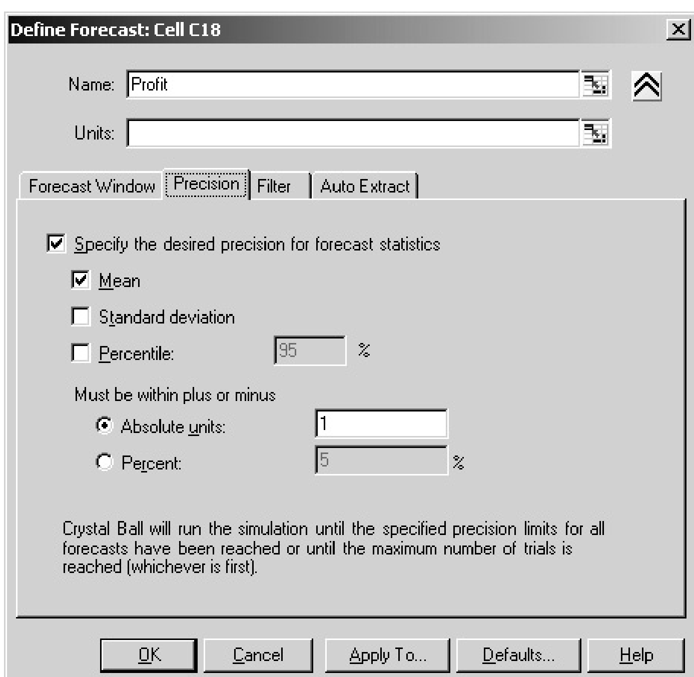


nes. Es una estimación de la desviación estándar del promedio de la muestra, lo que señala que la mayor parte del tiempo el promedio de la muestra tiene un error de una desviación estándar medida desde la media verdadera. En otras palabras, es probable que ésta esté desviada de la media de la muestra en una cantidad superior al error estándar de la media, pero la mayor parte del tiempo (aproximadamente 68%), no se desviará más allá de esta cifra. Por lo tanto, el intervalo desde \$45.94 – \$0.62 = \$45.32 hasta \$45.94 + \$0.62 = \$46.56 es un *intervalo de confianza* de 68% de la media verdadera. En forma similar, se puede obtener un intervalo de confianza más grande mediante el uso de un múltiplo apropiado del error estándar de la media para restarlo y después sumarlo a la media de la muestra. Por ejemplo, el múltiplo apropiado de un intervalo de confianza de 95% es 1.965, por lo que dicho intervalo de confianza va desde \$45.94 – 1.965(\$0.62) = \$44.72 hasta \$45.94 + 1.965(\$0.62) = \$47.16. (Este múltiplo de 1.965 cambiará un poco si el número de iteraciones es diferente a 500.) Por lo tanto, es muy probable que la media verdadera esté en algún lugar entre \$44.72 y \$47.16.

Si se requiere mayor precisión, por lo general el error estándar medio se puede reducir si se incrementa el número de iteraciones en la corrida de simulación. Sin embargo, la reducción tiende a ser pequeña aunque el número de iteraciones se incremente en forma sustancial. Por ejemplo, para reducir el error estándar medio a la mitad se requiere aproximadamente cuadriplicar el número de iteraciones. Por lo tanto, se puede requerir un número sorprendentemente grande de iteraciones para obtener el grado deseado de precisión.

Como el número de iteraciones necesario para obtener el grado de exactitud deseado no se puede predecir muy bien antes de la corrida de la simulación, la tentación es especificar un número muy grande de iteraciones. Este número puede resultar muchas veces más grande de lo necesario, lo que ocasiona una corrida en computadora excesivamente larga. Por fortuna, Crystal Ball tiene un método especial de control de precisión para detener la simulación tan pronto como se alcanza la precisión deseada. Este método se activa cuando se elige la opción (detenerse si se alcanza la precisión deseada, "*Stop when precision control limits has been reached*") en la caja de diálogo de iteraciones dentro de las preferencias de la corrida que se muestra en la figura 20.11. La precisión especificada se introduce en el cuadro de diálogo de definición del pronóstico "Define Forecast" que se despliega en la figura 20.15. (Este cuadro de diálogo surge al hacer clic sobre el botón "More"  en el cuadro de diálogo que se muestra en la figura 20.10.) El lado derecho inferior

FIGURA 20.15 Este cuadro de diálogo Define Forecast expandido se usa para especificar la precisión que se desea en la corrida de simulación de Freddie.



de la figura 20.15 indica que se aplica el control de precisión a la media (pero no a la desviación estándar o a un percentil específico). La corrida preferente de la figura 20.11 indica que se utiliza un intervalo de confianza de 95%. La amplitud de la mitad del intervalo de confianza, medida desde su punto medio hasta cualquiera de sus extremos, se considera la precisión que se ha alcanzado. La precisión deseada se puede especificar ya sea en términos absolutos (con las mismas unidades del intervalo de confianza), o en términos relativos (expresados como un porcentaje del punto medio del intervalo de confianza).

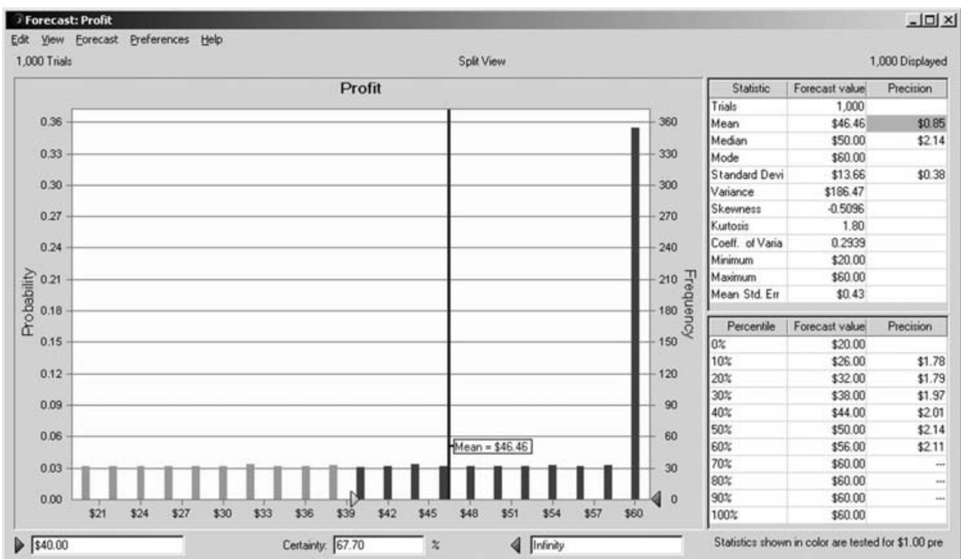
La figura 20.15 indica que la decisión se realizó para especificar la precisión deseada en términos absolutos como 1 dólar. Noventa y cinco por ciento del intervalo de confianza de la media después de 500 pruebas fue de $45.94 más o menos $1.22, por lo que $1.22 es la precisión que se logró después de estas pruebas. Crystal Ball también calculó el intervalo de confianza (así como la precisión actual) de forma periódica para verificar si el valor de la precisión actual estaba por debajo de $1, en cuyo caso se detendría la corrida. Sin embargo, esto nunca sucedió, por lo que Crystal Ball dejó que corriera la simulación hasta que se llegara al número máximo de pruebas (500).

Con el fin de obtener la precisión deseada, es necesario reiniciar la simulación con el fin de generar pruebas adicionales. Esto se llevó a cabo ingresando un número grande (como, por ejemplo, 5 000) como número máximo de pruebas (que incluía las 500 ya obtenidas) en el cuadro de diálogo Run Preferences (que se muestra en la figura 20.11) y después haciendo clic en el botón  Start Simulation. La figura 20.16 muestra los resultados que se obtuvieron después de llevar a cabo esta acción. El primer renglón indica que la precisión deseada se obtuvo sólo después de 500 pruebas adicionales, para un total de 1 000 pruebas. (El valor por omisión para la frecuencia de verificación de la precisión es de cada 500 pruebas, por lo que la precisión de $1 en realidad se logró en algún punto entre 500 y 1 000 pruebas.) Debido a las pruebas adicionales, han cambiado ligeramente algunas de las estadísticas con respecto a las que se proporcionaron en la figura 20.12. Por ejemplo, la mejor estimación de la media es ahora de $46.46, con una precisión de $0.85. Por lo tanto, es muy probable (con 95% de confianza) que el valor verdadero de la media esté dentro de $0.85 de $46.46.

$$\text{Intervalo de confianza de 95\%: } \$45.61 \leq \text{Media} \leq \$47.31$$

También se da la precisión de las estimaciones actuales de la mediana y la desviación estándar, así como de las estimaciones de los percentiles que se presentaron en la tabla correspondiente. Por lo tanto, también se puede calcular un intervalo de confianza de 95% de cada una de estas cantidades al sumar o restar su precisión a su estimación.

■ **FIGURA 20.16**
Resultados que se obtuvieron después de continuar con la corrida de simulación de Freddie hasta que se alcanza la precisión especificada en la figura 20.15.



Aplicación de la herramienta tabla de decisión

Los resultados que se presentaron en las figuras 20.12 y 20.16 fueron de una corrida de simulación que fijó en 60 la cantidad de ejemplares del *Financial Journal* que Freddie debía ordenar (como se indica en la celda C9 de la hoja de cálculo de la figura 20.7). Freddie decidió probar primero con esta cantidad porque parecía proporcionar un equilibrio razonable entre la capacidad de satisfacer por completo la demanda durante muchos días (alrededor de dos tercios) y la de no quedarse a menudo con muchos ejemplares sin vender en esos días. Sin embargo, los resultados que se obtuvieron no revelan si 60 es la cantidad por ordenar *óptima* que maximiza su ganancia diaria promedio. Se necesitan muchas más corridas de simulación con otras cantidades por ordenar para determinar (o al menos estimar) el valor óptimo.

Por fortuna, Crystal Ball proporciona una función especial llamada **herramienta tabla de decisión** (*Decision Table tool*) que aplica de manera sistemática la simulación para identificar al menos una aproximación a una solución óptima para los problemas que tienen sólo dos variables de decisión. El problema de Freddie tiene una sola variable de decisión, CantidadAOrdenar (C9), según el modelo en hoja de cálculo de la figura 20.7, por lo que a continuación se aplicará esta herramienta.

Un enfoque intuitivo para buscar una solución óptima sería utilizar prueba y error. Probar diferentes valores para las variables de decisión, correr una simulación con cada uno de ellos y observar cuál proporciona la mejor estimación de las medidas de desempeño elegidas. Esto es lo que hace la tabla de decisión, pero lo ejecuta de un modo sistemático. Sus cuadros de diálogo permiten especificar con rapidez lo que se desea hacer. Luego, después de hacer clic sobre un botón, corren todas las simulaciones deseadas y los resultados se despliegan con prontitud en la tabla de decisión. Si se desea, también pueden verse algunas gráficas, entre ellas la *gráfica de tendencia*, que proporcionan detalles adicionales acerca de los resultados.

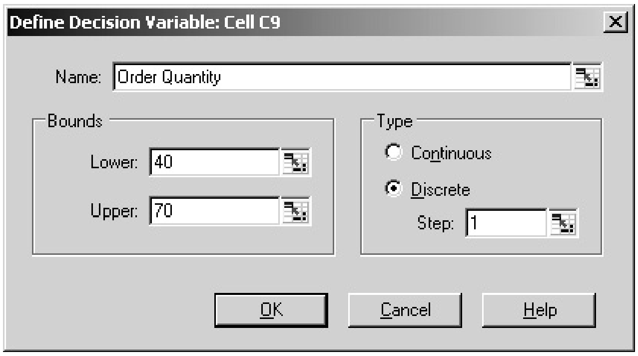
Para quienes han utilizado antes una tabla de datos de Excel o la tabla de Solver que se incluye en el OR Courseware para realizar análisis de sensibilidad en forma sistemática, la tabla de decisión funciona casi del mismo modo. En particular, la configuración de una tabla de decisión con una o dos variables de decisión es similar a la de una tabla de Solver con una o dos dimensiones (que se presentó en la sección 6.8). En una tabla de decisión, el máximo número de variables de decisión que se puede variar de manera simultánea es dos.

Como el número de ejemplares que los clientes de Freddie quieren comprar cambia ampliamente día con día (a cualquier punto entre 40 y 70 ejemplares), podría parecer adecuado comenzar por probar una muestra de posibles cantidades por ordenar, por ejemplo: 40, 45, 50, 55, 60, 65 y 70. Para hacer esto con la tabla de decisión, el primer paso es definir la variable de decisión que se investiga, CantidadAOrdenar (C9) en la figura 20.7, mediante el uso del siguiente procedimiento.

Procedimiento para definir una variable de decisión

1. Seleccione la celda que contiene la variable de decisión al hacer un clic sobre ella.
2. Si la celda todavía no contiene un valor, introduzca *cualquier* número en la celda.
3. Haga clic sobre el botón Define Decision  en la pestaña o en la barra de herramientas de Crystal Ball que trae el cuadro de diálogo para definir la variable de decisión (como se muestra en la figura 20.17 para el problema de Freddie).
4. Introduzca los límites inferior y superior del rango de valores que se simulará para la variable de decisión.
5. Haga clic sobre "Continuous" o "Discrete" para definir si la variable de decisión es continua o discreta.
6. Si en el paso 5 se selecciona la variable discreta, debe utilizarse el cuadro de paso ("Step") para especificar la diferencia entre valores posibles sucesivos (no sólo los que serán simulados) de la variable de decisión. (El valor predeterminado es 1.)
7. Haga clic sobre OK.

En la figura 20.17 se muestra la aplicación de este procedimiento al problema de Freddie. Como las simulaciones correrán para cantidades por ordenar que van desde 40 hasta 70, estos límites para el intervalo se han introducido a la izquierda. La cantidad por ordenar puede tener cualquier valor entero dentro de este intervalo, lo cual se indica a la derecha.

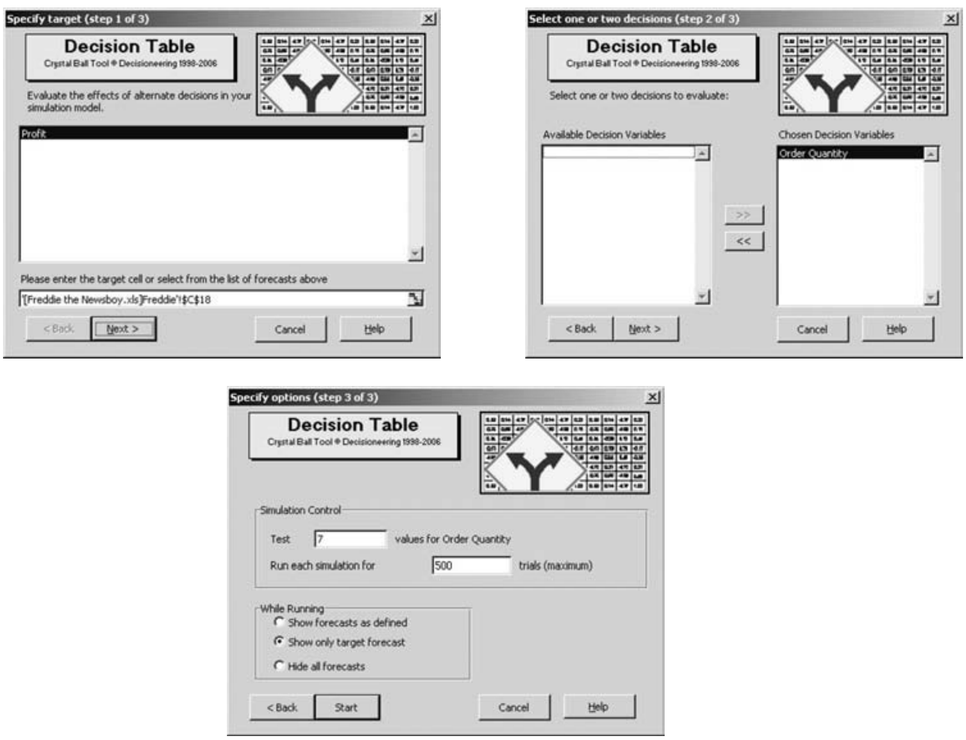


■ **FIGURA 20.17**
Este cuadro de diálogo para definir la variable de decisión especifica las características de la variable de decisión CantidadAOrdenar (C9) del modelo de simulación de la figura 20.7 en el ejemplo de Freddie, el voceador.

Ahora ya es posible elegir "Decision Table" del menú de herramientas de Crystal Ball. Después surge la secuencia de tres cuadros de diálogo que se muestra en la figura 20.18.

El primer cuadro de diálogo "Step" se utiliza para elegir una de las celdas de pronóstico enumeradas ahí como la celda objetivo de la tabla de decisión. El modelo en hoja de cálculo de Freddie en la figura 20.7 sólo tiene una celda de pronóstico, Ganancia (C18), por lo cual se selecciona ésta y se hace clic en el botón "Next".

En un inicio, el lado izquierdo del segundo cuadro de diálogo "Step" incluye una lista de todas las celdas que se han definido como variables de decisión. En el problema de Freddie, ésta consiste en la única variable de decisión, CantidadAOrdenar (C9). El propósito de este cuadro de diálogo es elegir cuál de las dos variables de decisión se elegirá para ser variada. Ello se hace mediante la selección de dichas variables de decisión del lado izquierdo, para después hacer clic sobre el botón con doble flecha a la derecha (>>) que se encuentra entre los dos cuadros, lo que lleva las variables de decisión hacia el lado derecho. En la figura 20.19 se muestra el resultado de ejecutar las acciones anteriores con la variable de decisión de Freddie.



■ **FIGURA 20.18**
Para preparar la generación de una tabla de decisión (Decision Table), estos tres cuadros de diálogo especifican 1) cuál celda de pronóstico será la celda objetivo, 2) cuál de las dos variables de decisión será variada y 3) las opciones de corrida. Las selecciones que se hacen aquí son para el ejemplo de Freddie, el voceador.

El tercer cuadro de diálogo “Step” se usa para especificar las opciones de la tabla de decisión. El primer cuadro de entrada registra el número de valores de la variable de decisión para los cuales se correrá las simulaciones. Entonces, Crystal Ball distribuye los valores de manera proporcional a lo largo del intervalo de valores especificado en el cuadro de diálogo para definir la variable de decisión (figura 20.17). En el problema de Freddie, el rango de valores es de 40 a 70, por lo que, al introducir 7 en el tercer cuadro de diálogo “Step”, se elige 40, 45, 50, 55, 60, 65 y 70 como los siete valores de la cantidad por ordenar para los cuales se correrá las simulaciones. Después de seleccionar el tamaño de corrida de cada simulación y especificar lo que se desea ver mientras corre la simulación, el último paso es hacer clic sobre el botón de inicio (“Start”).

Después de que Crystal Ball corre las simulaciones, se crea la tabla de decisión en una hoja de cálculo nueva, como se muestra en la figura 20.19. Para cada una de las cantidades por ordenar que aparecen en la parte superior, la fila 2 proporciona la media de los valores de la celda objetivo, Ganancia (C18), que se obtuvo en todas las iteraciones de esa corrida de simulación. Las celdas D2:F2 revelan que una cantidad ordenada de 55 produce la mayor ganancia media de $47.49, mientras que las cantidades ordenadas de 50 y 60, en esencia, empatan en segundo lugar.

La fuerte caída de las ganancias medias en ambos lados de estas cantidades por ordenar garantizan de manera virtual que la cantidad óptima está entre 50 y 60 (y probablemente cerca de 55). Para tener una mayor precisión, el siguiente paso lógico sería generar otra tabla de decisión que considere todos los valores enteros entre 50 y 60. En el problema 20.6-6 se le pide hacer esta tarea. (El tercer complemento de este capítulo en el sitio en internet de este libro usará el módulo OptQuest de Crystal Ball para especificar de otra manera la cantidad a ordenar óptima.)

La esquina superior izquierda de la tabla de decisión proporciona tres opciones para obtener información más detallada acerca de los resultados de las corridas de simulación para las celdas seleccionadas. Una opción es ver la gráfica de pronóstico que interesa como una gráfica de frecuencia o una gráfica acumulada, al seleccionar una celda de pronóstico en el renglón 2, para después hacer clic sobre el botón Forecast Charts. Otra opción es ver los resultados de dos o más corridas de simulación juntas. Esto se hace al seleccionar un conjunto de celdas de pronóstico, por ejemplo, las celdas E2:F2 de la figura 20.19, y después al hacer clic en el botón para superponer gráficas “Overlay Chart”. La gráfica superpuesta que resulta se muestra en la figura 20.20. Las líneas oscuras muestran la gráfica de frecuencia para la celda E2 (una cantidad a ordenar de 55) mientras que la líneas claras hacen lo mismo para la celda F2 (una cantidad a ordenar de 60), con lo cual los resultados para estos dos casos se pueden comparar uno junto al otro. (En un monitor a color se pueden ver los diferentes colores que se utilizan para distinguir entre los diferentes casos.)

La tercera opción es seleccionar todas las celdas de pronóstico que interesan (celdas B2:H2 en la figura 20.19) y después hacer clic en el botón para una gráfica de tendencia “Trend Chart”. Este procedimiento genera una gráfica interesante, llamada *gráfica de tendencia*, la cual se muestra en la figura 20.21. Los puntos clave a lo largo del eje horizontal son las siete líneas verticales que corresponden a los siete casos (cantidades a ordenar de 40, 45, . . . , 70) para las cuales se corrieron las simulaciones. El eje vertical da los valores de ganancia que se obtuvo en las iteraciones de estas corridas de simulación. Las bandas de la gráfica resumen la información acerca de la distribución de frecuencia de los valores de ganancia de cada corrida de la simulación. (En un monitor a color, las bandas aparecen en color azul claro para la banda central, rojo para el par de bandas adyacentes, verde para el siguiente par y azul oscuro para el par de bandas exterior.) Estas bandas están cen-

■ FIGURA 20.19
Tabla de decisión para el problema de Freddie.

	A	B	C	D	E	F	G	H
1	Trend Chart Overlay Chart Forecast Charts	Order Quantity (40)	Order Quantity (45)	Order Quantity (50)	Order Quantity (55)	Order Quantity (60)	Order Quantity (65)	Order Quantity (70)
2		$40.00	$44.17	$46.66	$47.49	$46.64	$44.14	$39.97
3		1	2	3	4	5	6	7

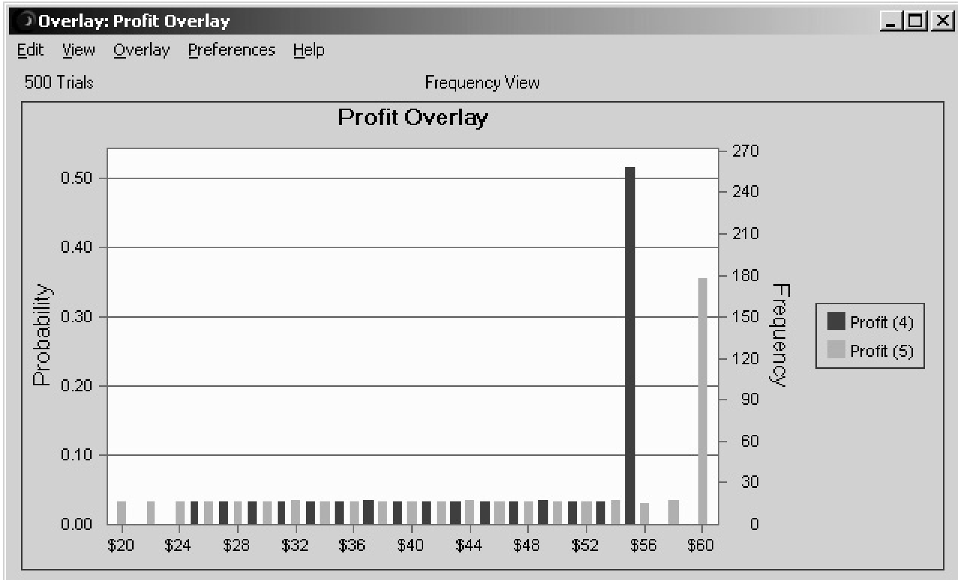
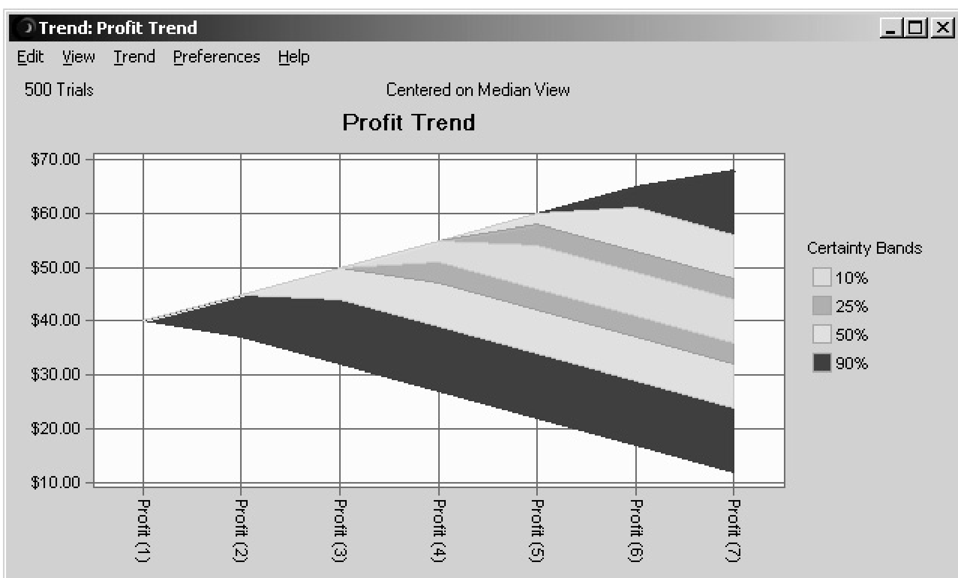


FIGURA 20.20
Gráfica sobrepuesta (Overlay) que compara las distribuciones de frecuencia de las cantidades a ordenar de 55 y 60 en el problema de Freddie.

tradas en las *medianas* de las distribuciones de frecuencia. En otras palabras, el centro de la banda de en medio (la más clara) proporciona un valor de ganancia tal que la mitad de las iteraciones da un valor más grande y la otra mitad un valor más pequeño. Esta banda central contiene 10% de los valores de ganancia que se encuentran en medio de la distribución (entonces, tiene 45% de los valores a cada lado). De manera similar, las tres bandas centrales contienen 25% de los valores de ganancia que están enmedio de la distribución, las cinco bandas de en medio contienen 50% de los valores, y todas las siete bandas contienen 90% de los valores de ganancia. (Estos porcentajes se enumeran a la derecha de la gráfica de tendencia.) De esta forma, 5% de los valores generados en las iteraciones de cada corrida de simulación cae por encima de la banda superior y otro 5% cae por debajo de la banda inferior.

La gráfica de tendencia recibe este nombre porque muestra las tendencias en forma gráfica a medida que se incrementa el valor de la variable de decisión (en este caso la cantidad por ordenar). Por ejemplo, en la figura 20.21 considere la banda central (que está oculta en



■ **FIGURA 20.21**
Gráfica de tendencia (Trend) que muestra la tendencia en el intervalo de varias partes de la distribución de frecuencia conforme se incrementa la cantidad a ordenar en el problema de Freddie.

la parte delgada de la gráfica de la izquierda). Al ir de la tercera cantidad a ordenar (50) a la cuarta (55), la banda central tiende hacia arriba, pero de ahí en adelante tiende hacia abajo. Por lo tanto, el valor mediano de los valores de ganancia generados en la simulación respectiva se incrementa a medida que aumenta la cantidad por ordenar hasta que la mediana alcanza su pico en una cantidad de 55, después de lo cual tiende hacia abajo. En forma similar, la mayoría de las otras bandas también tienden hacia abajo cuando la cantidad por ordenar se incrementa por encima de 55. Estas tendencias sugieren que una cantidad por ordenar de 55 es particularmente atractiva en términos de su distribución de frecuencia completa y no sólo por su valor medio. El hecho de que la gráfica de tendencia se disperse conforme se mueve hacia la derecha proporciona la impresión de que la variabilidad de los valores de ganancia se incrementa a medida que la cantidad a ordenar aumenta. Aunque las cantidades a ordenar más grandes dan alguna oportunidad de ganancias particularmente altas en ciertos días, también pueden conducir a una ganancia inusualmente baja durante algunos días. El perfil del riesgo puede ser relevante para Freddie si está preocupado por la variabilidad de sus ganancias diarias.

Si desea leer más acerca de cómo realizar simulaciones en hojas de cálculo con Crystal Ball, en el capítulo 28 del sitio en internet de este libro se proporcionan algunos ejemplos adicionales y mayores detalles. Estos ejemplos incluyen aplicaciones para licitación de contratos, administración de proyectos, administración de flujos de efectivo, análisis de riesgo financiero y administración de utilidades.

20.7 CONCLUSIONES

La simulación es una herramienta que se emplea con amplitud para estimar el desempeño de sistemas estocásticos complejos cuando se quiere usar nuevos diseños o políticas de operación.

Este capítulo se dedicó al uso de la simulación para predecir el comportamiento de *estado estable* de sistemas cuyos estados cambian sólo en puntos discretos del tiempo. Sin embargo, al hacer que una serie de corridas de simulación comiencen con las *condiciones iniciales* preestablecidas, también se puede usar la simulación para describir el comportamiento *transitorio* de un sistema propuesto. Más aún, si se usan ecuaciones diferenciales, la simulación se puede aplicar a sistemas cuyos estados cambian en forma *continua* en el tiempo.

La simulación es una de las técnicas que más se emplean en investigación de operaciones debido a que es una herramienta flexible, poderosa e intuitiva. En cuestión de segundos o minutos, puede simular incluso años de operación de un sistema común mientras genera una serie de observaciones estadísticas sobre el desempeño del sistema en este periodo. Debido a su excepcional versatilidad, la simulación se ha aplicado en una amplia variedad de áreas. Más aún, sus horizontes continúan ampliándose debido a los grandes avances en el desarrollo de software de simulación, que incluye el que realiza simulaciones en hojas de cálculo.

Por otro lado, la simulación no debe verse como una panacea al estudiar sistemas estocásticos. Cuando se pueden aplicar los métodos analíticos (como los de los capítulos 15 a 19) se obtienen algunas ventajas significativas. La simulación es una técnica por naturaleza imprecisa. Proporciona sólo *estimaciones estadísticas* y no resultados exactos y *compara alternativas* más que generar una solución óptima (a menos que se use un paquete de software especial como el OptQuest, el cual se describe en el tercer complemento de este capítulo en el sitio en internet de este libro). A pesar de los grandes avances en el software, todavía puede ser una manera *lenta* y *costosa* de estudiar sistemas estocásticos complejos. Por lo general requiere una gran cantidad de tiempo y gasto para el análisis y la programación, además de un tiempo considerable para las corridas en computadora. Los modelos de simulación tienden a complicarse, por lo que el número de casos que se pueden correr y la exactitud de los resultados que se obtienen con frecuencia son inadecuados. Por último, la simulación conduce sólo a *datos numéricos* sobre el comportamiento del sistema, no proporciona una visión adicional respecto de las relaciones de causa y efecto dentro del sistema, excepto por los indicios que pueden extraerse a partir de estos números (y del análisis que se requiere para construir el modelo). Por lo tanto, es muy costoso llevar a cabo un análisis de sensibilidad de los valores de los parámetros supuestos por el modelo. La única forma posible sería efectuar una serie de corridas de simulación con diferentes valores de los parámetros que tenderían a proporcionar muy poca información a un costo relativamente alto.

Por todas estas razones, los métodos analíticos (cuando están disponibles) y la simulación tienen papeles complementarios importantes cuando se estudian sistemas estocásticos. Un método analítico es adecuado para hacer al menos un análisis preliminar, a fin de examinar las relaciones causa-efecto, para cierta optimización global y para realizar análisis de sensibilidad. Cuando el modelo matemático del método analítico no capta todas las características importantes de los sistemas estocásticos, la simulación puede incorporar a todas ellas y obtener información detallada acerca de las medidas de desempeño de algunos candidatos importantes para la configuración final del sistema.

La simulación proporciona una manera de *experimentar* con las políticas o sistemas propuestos sin tener que hacer cambios en el sistema real. El diseño de estos experimentos debe estar bien fundamentado en la teoría estadística. Con frecuencia se necesitan corridas de simulación sorprendentemente largas para obtener resultados *estadísticamente significativos*. No obstante, las *técnicas de reducción de varianza* (descritas en el primer complemento de este capítulo en el sitio en internet de este libro) pueden ser una gran ayuda para reducir la longitud necesaria de las corridas.

Cuando se aplican los procedimientos tradicionales de la estimación estadística, surgen varios problemas tácticos en los experimentos simulados. Estos problemas incluyen la indicación de las *condiciones iniciales* apropiadas, la determinación de la longitud del *periodo de calentamiento* para alcanzar en esencia la condición de estado estable y el manejo de las observaciones *estadísticamente dependientes*. Estos problemas se pueden eliminar con el *método regenerativo* de análisis estadístico (descrito en el segundo complemento de este capítulo en el sitio en internet de este libro). Sin embargo, existen algunas restricciones sobre el momento en que se puede aplicar este método.

No hay duda de que la simulación ocupa un lugar importante en la teoría y práctica de la IO. Es una herramienta invaluable para usarla en aquellos problemas en los que las técnicas analíticas son inadecuadas y, en realidad, su uso es cada vez mayor.

REFERENCIAS SELECCIONADAS

1. Argon, N. T. y S. Andradóttir: "Replicated Batch Means for Steady-State Simulations", en *Naval Research Logistics*, **53**(6): 508-524, septiembre de 2006.
2. Asmussen, S. y P. W. Glynn: *Stochastic Simulation*, Springer, Nueva York, 2007.
3. Banks, J., J. S. Carson, II, B. L. Nelson y D. M. Nicol: *Discrete-Event System Simulation*, 4a. ed., Prentice-Hall, Upper Saddle River, NJ, 2005.
4. Branke, J., S. E. Chick y C. Schmidt: "Selecting a Selection Procedure", en *Management Science*, **53**(12): 1916-1932, diciembre de 2007.
5. Del Castillo, E.: *Process Optimization: A Statistical Approach*, Springer, Nueva York, 2007.
6. Fishman, G. S.: *Discrete-Event Simulation: Modeling, Programming and Analysis*, Springer, Nueva York, 2001.
7. Fu, M. C.: "Optimization for Simulation: Theory vs. Practice", en *INFORMS Journal on Computing*, **14**: 192-215, 2002.
8. Kleijnen, J. P. C.: *Design and Analysis of Simulation Experiments*, Springer, Nueva York, 2008.
9. Kleijnen, J. P. C., S. M. Sanchez, T. W. Lucas y T. M. Cioppa: "State-of-the-Art Review: A User's Guide to the Brave New World of Designing Simulation Experiments", en *INFORMS Journal on Computing*, 17(3): 263-289, verano de 2005.
10. Law, A. M. y W. D. Kelton: *Simulation Modeling and Analysis*, 3a. ed., McGraw-Hill, Nueva York, 2000.
11. Nance, R. E. y R. G. Sargent: "Perspectives on the Evolution of Simulation", en *Operations Research*, **50**(1): 161-172, enero-febrero de 2002.
12. Swain, J.: "Software Survey: New Frontiers in Simulation", en *OR/MS Today*, **34**(5): 32-43, octubre de 2007.
13. Tekin, E. e I. Sabuncuoglu: "Simulation Optimization: A Comprehensive Review on Theory and Applications", en *IIE Transactions*, **36**(11): 1067-1081, noviembre de 2004.
14. Whitt, W: "Planning Queueing Simulations", en *Management Science*, **35**(11): 1341-1366, noviembre de 1989.

Algunas aplicaciones de simulación que han ganado premios:

(En el sitio en internet de este libro —www.mhhe.com/hillier— se proporciona una liga hacia estos artículos.)

A1. Alden, J. M., L. D. Burns, T. Costy, R. D. Hutton, C. A. Jackson, D. S. Kim, K. A. Kohls, J. H. Owen, M. A. Turnquist y D. J. Vander Veen: "General Motors Increases Its Production Throughput", en *Interfaces*, 36(1): 6-25, enero-febrero, 2006.

A2. Barabba, V., C. Huber, F. Cooke, N. Pudar, J. Smith y M. Paich: "A Multimethod Approach for Creating New Business Models: The General Motors OnStar Project", en *Interfaces*, **32**(1): 20-34, enero-febrero de 2002.

A3. Beis, D. A., P. Loucopoulos, Y. Pyrgiotis y K. G. Zografos: "PLATO Helps Athens Win Gold: Olympic Games Knowledge Modeling for Organizational Change and Resource Management", en *Interfaces*, **36**(1): 26-42, enero-febrero de 2006.

A4. Brinkley, P. A., D. Stepto, K. R. Haag, J. Folger, K. Wang, K. Liou y W. D. Carr: "Nortel Redefines Factory Information Technology: An OR-Driven Approach", en *Interfaces*, **28**(1): 37-52, enero-febrero de 1998.

A5. Cebry, M. E., A. H. DeSilva y F. J. DiLisio: "Management Science in Automating Postal Operations: Facility and equipment Planing in the United States Postal Service", en *Interfaces* **22**(1): 110-130, enero-febrero de 1992.

A6. Duffy, T., M. Hatzakis, W. Hsu, R. Labe, B. Liao, X. Luo, J. Oh, A. Setya y L. Yang: "Merrill Lynch Improves Liquidity Risk Management for Revolving Credit Lines", en *Interfaces*, **35**(5): 353-369, septiembre-octubre de 2005.

A7. Hueter, J. y W. Swart: "An Integrated Labor-Management System for Taco Bell", en *Interfaces*, **28**(1): 75-91, enero-febrero de 1998.

A8. Larson, R. C., M. F. Cahn y M. C. Shell: "Improving the New York City Arrest-to-Arraignment System", en *Interfaces*, **23**(1): 76-96, enero-febrero de 1993.

A9. Mulvey, J. M., G. Gould y C. Morgan: "An Asset and Liability Management System for Towers Perrin-Tillinghast", en *Interfaces*, **30**(1): 96-114, enero-febrero de 2000.

A10. Pfeil, G., R. Holcomb, C. T. Muir y S. Taj: "Visteon's Sterling Plant Uses Simulation-Based Decision Support in Training, Operations and Planning", en *Interfaces*, **30**(1): 115-133, enero-febrero de 2000.

AYUDAS DE APRENDIZAJE PARA ESTE CAPÍTULO EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

Worked Examples

Ejemplos para el capítulo 20

Ejemplos de demostración en el OR Tutor:

Simulación de un sistema de colas básico
Simulación de un sistema de colas con prioridades

Rutina automática en el IOR Tutorial:

Animación de un sistema de colas (*Animation of a Queueing System*)

Rutinas interactivas en el IOR Tutorial:

Introducción de un problema de colas (*Enter Queueing Problem*)
Simulación interactiva de un problema de colas (*Interactively Simulate Queueing Problem*)

Archivos de Excel "Ch. 20—Simulation":

Ejemplos en hojas de cálculo
Simulador de colas (*Queueing Simulator*)

Complementos de Excel:

RiskSim (versión académica)

Glosario del capítulo 20

Complementos a este capítulo

Técnicas de reducción de la varianza (*Variance-Reducing Techniques*)
Método regenerativo de análisis estadístico (*Regenerative Method of Statistical Analysis*)
Optimización con OptQuest

Vea el apéndice 1 para documentación del software.

PROBLEMAS

Los símbolos a la izquierda de algunos problemas (o de sus incisos) significan lo siguiente:

D: Los ejemplos de demostración de este capítulo pueden ser útiles.
I: Se sugiere usar las rutinas interactivas indicadas en las ayudas de aprendizaje (la impresión registra su trabajo).
E: Use Excel.
A: Use alguno de los complementos de Excel para simulación, como RiskSim o Crystal Ball.
Q: Use el simulador de colas.
R: Utilice números aleatorios uniformes de *tres dígitos* (0.096, 0.569, etc.) que se obtuvieron con los dígitos aleatorios consecutivos de la tabla 20.3; comience al inicio del primer renglón, para contestar cada inciso del problema.

20.1-1.* Utilice los números aleatorios uniformes de las celdas C13:C18 de la figura 20.1 para generar seis observaciones aleatorias de las siguientes situaciones.

a) El lanzamiento de una moneda no cargada.
b) Un pitcher de beisbol que lanza un strike 60% de las veces y una bola 40% de las veces.
c) El color de la luz del semáforo que encuentra un automóvil que llega al azar, si 40% del tiempo está en verde, 10% en amarillo y 50% en rojo.

20.1-2. El clima se puede considerar un sistema estocástico, porque evoluciona de una manera probabilística de un día para otro. Suponga que en cierto lugar este comportamiento probabilístico satisface la siguiente descripción:

La probabilidad de lluvia para mañana es de 0.6 si hoy llueve. La probabilidad de un día despejado (sin lluvia) para mañana es de 0.8 si hoy está despejado.

a) Emplee los números aleatorios uniformes de las celdas C17:C26 de la figura 20.1 para simular el comportamiento del clima durante 10 días; comience con un día que sigue a uno despejado.
E ***b***) Ahora use una computadora con números aleatorios uniformes generados por Excel para realizar la simulación del inciso *a*) en una hoja de cálculo.

20.1-3. Jessica Williams, gerente de Kitchen Appliances de la tienda Midtown, cree que sus niveles de inventario de estufas son más altos de lo necesario. Antes de corregir la política de inventarios, registra el número vendido cada día durante un periodo de 25 días, como se resume a continuación.

Cantidad vendida	2	3	4	5	6
Número de días	4	7	8	5	1

a) Use estos datos para estimar la distribución de probabilidad de las ventas diarias.
b) Calcule la media de la distribución del inciso *a*).
c) Describa cómo se puede usar números aleatorios uniformes para simular las ventas diarias.
d) Use los números aleatorios uniformes 0.4476, 0.9713 y 0.0629 para simular las ventas diarias durante 3 días. Compare el promedio con la media que obtuvo en *b*).
E ***e***) Formule un modelo en hoja de cálculo para simular las ventas diarias. Realice 300 réplicas y obtenga el promedio de ventas de los 300 días simulados.

20.1-4. La William Graham Entertainment Company abrirá una nueva taquilla donde los clientes puedan ir a comprar boletos por adelantado para los muchos eventos que se llevan a cabo en el área. Se ha empleado simulación para analizar si coloca uno o dos dependientes en la taquilla.

Según la simulación, al comenzar el día en la taquilla, el primer cliente llega 5 minutos después de abrir y los tiempos entre llegadas de los siguientes 4 clientes (en orden) son 3, 9, 1 y 4 minutos, después de lo cual hay un intervalo largo hasta la llegada del siguiente. Los tiempos de servicio son (en orden) 8, 6, 2, 4 y 7 minutos.

a) En el caso de la alternativa de un solo dependiente, grafique la evolución del número de clientes en la taquilla en este periodo.
b) Use la gráfica que construyó para estimar las medidas usuales de desempeño —L, L_q, W, W_q y las P_n (definidas en la sección 17.2)— de este sistema de colas.
c) Repita el inciso *a*) para la alternativa de dos dependientes.
d) Repita el inciso *b*) para la alternativa de dos dependientes.

20.1-5. Considere el modelo de colas *M*/*M*/1 de la sección 17.6 y del ejemplo 2 de la sección 20.1. Suponga que la tasa media de llegadas es de 10 por hora, mientras que la tasa media de servicio es de 12 por hora; estime, con simulación, el tiempo de espera esperado antes de comenzar el servicio.

R ***a***) Inicie con el sistema vacío; utilice el método de incrementos por evento para realizar una simulación manual hasta que ocurran dos terminaciones de servicio.
R ***b***) A partir de un sistema vacío, use el método de incrementos de tiempo fijo (con 2 minutos como unidad de tiempo) para

realizar una simulación manual hasta que hayan ocurrido dos terminaciones de servicio.

D,I ***c***) Utilice la rutina interactiva para simulación del IOR Tutorial (la cual incorpora incrementos por evento) para ejecutar una corrida de simulación de 20 terminaciones de servicio.

Q ***d***) Utilice el simulador de colas para hacer una corrida de simulación con 10 000 llegadas de clientes.

E ***e***) Utilice la plantilla de Excel del capítulo 17 de este modelo para obtener las medidas de desempeño usuales del sistema de colas. Después compare estos resultados exactos con las estimaciones puntuales correspondientes y el intervalo de 95% de confianza que se obtuvo con la corrida de simulación en *d*). Identifique cualquier medida cuyo resultado exacto esté fuera de este intervalo.

20.1-6. La Rustbelt Manufacturing Company emplea una cuadrilla de mantenimiento para reparar las máquinas cuando se necesita. La gerencia desea realizar un estudio de simulación para analizar cuál debe ser el tamaño de la cuadrilla, donde los tamaños posibles son 2, 3 y 4. El tiempo que requiere la cuadrilla para reparar una máquina tiene distribución uniforme entre 0 y el doble de la media, donde la media depende del tamaño de la cuadrilla. Esta media es de 4 horas con 2 personas, 3 con 3 personas y 2 con 4. El tiempo entre descomposturas tiene distribución exponencial con media de 5 horas. La administración quiere que cuando se descomponga una máquina, su tiempo de espera promedio antes de iniciar la reparación no sea mayor de 3 horas. También desea que el tamaño de la cuadrilla no sea más grande del necesario para lograr este objetivo.

a) Desarrolle un modelo de simulación para el problema; describa los bloques básicos de construcción enumerados en la sección 20.1 según se apliquen a esta situación.

R ***b***) Considere una cuadrilla de tamaño 2. Comience sin máquinas descompuestas y use incrementos por evento para realizar una simulación a mano de 20 horas.

R ***c***) Repita el inciso *b*) con incrementos de tiempo fijo (con 1 hora como unidad de tiempo).

D,I ***d***) Use la rutina interactiva para simulación del IOR Tutorial (la cual incorpora incrementos por evento) para ejecutar una corrida de simulación de hasta 10 descomposturas para cada tamaño de cuadrilla.

Q ***e***) Use el simulador de colas; simule el sistema hasta 10 000 descomposturas para cada tamaño.

f) Con el modelo *M*/*G*/1 de la sección 17.7 obtenga el tiempo de espera promedio W_q en forma analítica según cada tamaño de cuadrilla. (Puede calcular W_q a mano o usar la plantilla del modelo del archivo de Excel del capítulo 17.) ¿Cuál debe ser el tamaño de la cuadrilla?

20.1-7. Al realizar una simulación de un sistema de colas de un solo servidor, el número de clientes en el sistema es 0 los primeros 10 minutos, 1 los siguientes 17, 2 los siguientes 24, 1 los siguientes 15, 2 los siguientes 16 y 1 los 18 que siguen. Después de este total de 100 minutos, el número de clientes es 0 de nuevo. Con base en estos resultados realice el siguiente análisis (con la notación de modelos de colas de la sección 17.2).

a) Grafique la evolución del número de clientes en el sistema durante estos 100 minutos.

b) Estime P_0, P_1, P_2, P_3.

c) Estime L y L_q.

d) Estime W y W_q.

20.1-8. Vea el primer ejemplo de demostración (*Simulating a Basic Queueing System*) del área de simulación del OR Tutor.

D,I ***a***) Introduzca este *mismo problema* en la rutina interactiva para simulación del IOR Tutorial. Ejecute una corrida interactiva de simulación de 20 minutos.

Q ***b***) Use el simulador de colas con 5 000 llegadas de clientes para estimar las medidas de desempeño usuales de este sistema con el plan actual de proporcionar 2 cajeros.

Q ***c***) Repita el inciso *b*) si se asignan tres cajeros.

Q ***d***) Ahora realice un análisis de sensibilidad para verificar el efecto si el nivel de negocios resulta aún más alto que el proyectado. En particular, suponga que el tiempo promedio entre llegadas resulta de sólo 0.9 minutos en lugar de 1 minuto. Evalúe las alternativas de dos y tres cajeros bajo este supuesto.

e) Suponga que *usted* es el gerente del banco. Use los resultados de simulación para tomar una decisión gerencial sobre el número de cajeros. Justifique su respuesta.

D,I **20.1-9.** Vea el segundo ejemplo de demostración de un sistema de colas con prioridades (*Simulating a Queueing System with Priorities*) en el área de simulación del OR Tutor. Introduzca este *mismo problema* en la rutina interactiva de simulación del IOR Tutorial. Ejecute una corrida interactiva de 20 minutos.

20.1-10.* Hugh's Repair Shop se especializa en autos alemanes y japoneses. El taller tiene dos mecánicos. Uno trabaja sólo en autos alemanes y el otro sólo en los japoneses. En cualquier caso, el tiempo que se requiere para reparar un auto tiene distribución exponencial con media de 0.2 días. El negocio ha crecido en forma estable, en especial el de los autos alemanes. Hugh piensa que para el próximo año, los autos alemanes llegarán de manera aleatoria con tasa media de 4 por día, y el tiempo entre llegadas tendrá distribución exponencial con media de 0.25 días. Cree que la tasa media de llegadas de los autos japoneses será de 2 por día, y los tiempos entre llegadas tendrán distribución exponencial con media de 0.5 días.

Para cualquier tipo de auto, Hugh quiere que el tiempo de espera promedio antes de la reparación no sea mayor de 0.5 días.

a) Formule un modelo de simulación para estimar cuál será el año próximo el tiempo de espera promedio hasta que termine la reparación de cualquier auto.

D,I ***b***) Considere sólo los autos alemanes; use la rutina interactiva de simulación del IOR Tutorial para simular un periodo de 10 llegadas de este tipo de autos.

Q ***c***) Utilice el simulador de colas para realizar esta simulación de autos alemanes para 10 000 llegadas.

Q ***d***) Repita el inciso *c*) para los autos japoneses.

D,I ***e***) Hugh piensa contratar otro mecánico especializado en autos alemanes para reparar dos de estos autos a la vez (sólo un mecánico trabaja en un auto). Repita el inciso *b*) para esta opción.

Q ***f***) Utilice el simulador de colas con 10 000 llegadas de autos alemanes para evaluar la opción del inciso *e*).

Q ***g***) Otra opción es capacitar a los dos mecánicos para que trabajen en cualquier auto. Esta alternativa aumentaría 10% el tiempo esperado de reparación, de 0.2 días a 0.22 días. Use el simulador de colas con 20 000 llegadas de cualquier tipo para evaluar esta opción.

h) Como las distribuciones de los tiempos entre llegadas y de servicio son exponenciales, se pueden usar los modelos *M*/*M*/1 y *M*/*M*/*s* de la sección 17.6 para realizar una evaluación analítica de las opciones anteriores. Determine *W*, el tiempo de espera

promedio hasta terminar la reparación, para cada caso en *c*), *d*), *f*) y *g*). (Puede calcular *W* a mano o usar la plantilla de Excel para el modelo *M*/*M*/*s* de los archivos de Excel para el capítulo 17.) En cada caso, compare la estimación de *W* que obtuvo mediante simulación en computadora con el valor analítico. ¿Qué dice esto del número de llegadas que debe incluirse en la simulación?

i) Con base en los resultados anteriores, ¿qué opción seleccionaría si fuera Hugh? ¿Por qué?

20.1-11. Vistaprint produce monitores e impresoras para computadoras. En el pasado, sólo algunos se inspeccionaban por muestreo, pero el nuevo plan es que todos se inspeccionen antes de salir. Con este plan, los monitores e impresoras se traerán a la estación de inspección uno a la vez cuando estén terminados. Para los monitores, el tiempo entre llegadas es uniforme entre 10 y 20 minutos. En el caso de las impresoras es constante de 15 minutos.

La estación de inspección tiene dos inspectores. Uno trabaja sólo con monitores y el otro sólo con impresoras. En ambos casos, el tiempo de inspección tiene distribución exponencial con media de 10 minutos.

Antes de iniciar el nuevo plan, la administración desea evaluar cuánto tardarán los monitores e impresoras en la estación de inspección.

a) Formule un modelo para realizar una simulación que estime los tiempos de espera promedio (antes y después de la inspección) tanto de los monitores como de las impresoras.

D,I ***b***) Considere sólo los monitores; use la rutina interactiva de simulación del IOR Tutorial para simular hasta 10 llegadas de monitores.

D,I ***c***) Repita el inciso *b*) para las impresoras.

Q ***d***) Utilice el simulador de colas para repetir los incisos *b*) y *c*) con 10 000 llegadas en cada caso.

Q ***e***) La administración estudia la opción de proporcionar a los inspectores un nuevo equipo de inspección que no cambiaría el tiempo esperado de inspección pero disminuiría la variabilidad de los tiempos. En particular, para cualquiera de los productos, el tiempo de inspección tendría distribución Erlang con media de 10 minutos y parámetro de forma $k = 4$. Use el simulador de colas para repetir el inciso *d*) con esta opción. Compare estos resultados con los que obtuvo en el inciso *d*).

20.2-1. Lea el artículo de referencia que describe el estudio de IO que se resume en el Recuadro de aplicación que se presentó en la sección 20.2. Describa de manera breve la forma en que se aplicó la simulación en este estudio. Después, elabore una lista de los beneficios financieros y de otro tipo que arrojó dicho estudio.

20.2-2. En la sección 20.2 se introdujeron cuatro aplicaciones reales de simulación descritas en las referencias seleccionadas A1 y A5. Seleccione una de ellas y lea el artículo correspondiente. Escriba un resumen de dos páginas de la aplicación y los beneficios que proporcionó.

20.3-1.* Use el método congruencial mixto para generar las siguientes sucesiones de números aleatorios.

a) Una sucesión de 10 números aleatorios enteros de *un dígito*, tal que $x_{n+1} \equiv (x_n + 3)$ (módulo 10) y $x_0 = 2$.

b) Una sucesión de 8 números aleatorios enteros entre 0 y 7, tal que $x_{n+1} \equiv (5x_n + 1)$ (módulo 8) y $x_0 = 1$.

c) Una sucesión de 5 números aleatorios enteros de *dos dígitos*, tal que $x_{n+1} \equiv (61x_n + 27)$ (módulo 100) y $x_0 = 10$.

20.3-2. Reconsidere el problema 20.3-1. Suponga ahora que se desea convertir estos números aleatorios enteros en números con distribución uniforme (aproximada). Para cada inciso, dé una fórmula para esta conversión que haga la aproximación tan cercana como sea posible.

20.3-3. Utilice el método congruencial mixto para generar una sucesión de 5 números aleatorios enteros de *dos dígitos*, tal que $x_{n+1} \equiv (11x_n + 23)$ (módulo 100) y $x_0 = 52$.

20.3-4. Use el método congruencial mixto para generar una sucesión de 3 números aleatorios enteros de *tres dígitos*, tal que $x_{n+1} \equiv (201x_n + 503)$ (módulo 1 000) y $x_0 = 485$.

20.3-5. Usted necesita generar cinco números aleatorios uniformes.

a) Para ello utilice el método congruencial mixto y genere una sucesión de 5 números aleatorios enteros entre 0 y 31 tal que $x_n+_1 \equiv (13x_n + 15)$(módulo 32) y $x_0 = 14$.

b) Convierta estos números aleatorios enteros en números aleatorios uniformes aproximándolos tan cerca como sea posible.

20.3-6. Se tiene un *generador congruencial multiplicativo* con $x_0 = 1$ y $x_{n+1} \equiv 7x_n$ (módulo 13) para $n = 0, 1, 2, \ldots$

a) Calcule x_n para $n = 1, 2, \ldots, 12$.

b) ¿Con qué frecuencia aparece cada número entre 1 y 12 en la sucesión generada en el inciso *a*)?

c) Sin realizar cálculos adicionales, compare $x_{13}, x_{14}, \ldots$, con $x_1, x_2, \ldots$

20.4-1. Reconsidere el juego lanzamiento de monedas de la sección 20.1 que se analizó con simulación en las figuras 20.1, 20.2 y 20.3.

a) Simule una jugada lanzando una moneda repetidas veces hasta que termine el juego. Registre sus resultados en el formato de las columnas *B*, *D*, *E*, *F* y *G* de la figura 20.1. ¿Cuánto habría ganado o perdido de haber sido un juego real?

E ***b***) Cambie el modelo de la hoja de cálculo de la figura 20.1 por medio de la función Excel VLOOKUP en lugar de la función "SI" para generar cada lanzamiento. Después realice una simulación de una jugada.

E ***c***) Use el modelo de hoja de cálculo corregido para generar una tabla con 14 réplicas como la de la figura 20.2.

E ***d***) Repita el inciso *c*) con 1 000 réplicas (como en la figura 20.3).

20.4-2.* Aplique el método de la transformación inversa como se indicó para generar tres observaciones de una distribución uniforme entre –10 y 40 con los siguientes números aleatorios uniformes: 0.0965, 0.5692, 0.6658.

a) Aplique este método de manera gráfica.

b) Aplique este método en forma algebraica.

c) Escriba la ecuación que usaría Excel para generar cada observación.

R **20.4-3.** Obtenga números aleatorios uniformes como se indicó al principio de los problemas; genere tres observaciones aleatorias a partir de cada una de las siguientes distribuciones de probabilidad:

a) La distribución uniforme de 25 a 75.

b) La distribución cuya función de densidad de probabilidad es

$$f(x) = \begin{cases} \frac{1}{4}(x+1)^3 & \text{si } -1 \le x \le 1 \\ 0 & \text{de otra manera.} \end{cases}$$

c) La distribución cuya función de densidad de probabilidad es

$$f(x) = \begin{cases} \dfrac{1}{200}(x - 40) & \text{si } 40 \leq x \leq 60 \\ 0 & \text{de otra manera.} \end{cases}$$

R **20.4-4.** Obtenga números aleatorios uniformes como se indicó, y genere tres observaciones aleatorias a partir de las siguientes distribuciones de probabilidad.

a) La variable aleatoria X tiene $P\{X = 0\} = \frac{1}{2}$. Si $X \neq 0$ tiene distribución uniforme entre –5 y 15.

b) La distribución cuya función de densidad de probabilidad es

$$f(x) = \begin{cases} x - 1 & \text{si } 1 \leq x \leq 2 \\ 3 - x & \text{si } 2 \leq x \leq 3. \end{cases}$$

c) La distribución geométrica con parámetro $p = \frac{1}{3}$, de manera que

$$P\{X = k\} = \begin{cases} \dfrac{1}{3}\left(\dfrac{2}{3}\right)^{k-1} & \text{si } k = 1, 2, \ldots \\ 0 & \text{de otra manera.} \end{cases}$$

20.4-5. Cada vez, una moneda no cargada se lanza tres veces. Las probabilidades respectivas de obtener 0, 1, 2 y 3 caras son $\frac{1}{8}$, $\frac{3}{8}$, $\frac{3}{8}$ y $\frac{1}{8}$, respectivamente. Entonces, con ocho grupos de tres lanzamientos, *en promedio*, un grupo dará 0 caras, tres grupos darán 1 cara, tres, 2 caras, y uno, 3 caras.

a) Use una moneda para lanzarla 24 veces en ocho grupos de tres y registre el número de grupos con 0, 1, 2 y 3 caras.

b) Obtenga números aleatorios uniformes como se indicó al principio de los problemas, simule los lanzamientos especificados en *a*) y registre la información.

E *c*) Formule un modelo en hoja de cálculo para simular tres lanzamientos de la moneda y registre el número de caras. Realice una réplica de esta simulación.

E *d*) Use esta hoja para calcular una tabla con 8 réplicas de la simulación. Compare esta distribución de frecuencia del número de caras con la distribución de probabilidad teórica de tres lanzamientos.

E *e*) Repita el inciso *d*) con 800 réplicas.

20.4-6.* Un juego de dados requiere que el jugador lance dos dados una o más veces hasta que se llegue a una decisión de si pierde o gana. Gana si la primera tirada suma 7 u 11, o alternativamente si la primera suma es 4, 5, 6, 8, 9 o 10 y sale la misma suma antes de que aparezca una suma de 7. Por el contrario, pierde si el resultado de la primera tirada suma 2, 3 o 12, o si la primera suma es 4, 5, 6, 8, 9 o 10 y aparece una suma de 7 antes de que la primera suma vuelva a salir.

E *a*) Formule un modelo en una hoja de cálculo para simular la tirada de dos dados. Realice una réplica.

E *b*) Realice 25 réplicas de esta simulación.

c) Analice estas 25 réplicas para determinar el número de veces que el jugador simulado habría ganado el juego de dados y el número de veces que lo habría perdido cuando cada jugada comienza con el siguiente lanzamiento después de que termina el juego anterior. Use esta información para calcular una estimación preliminar de la probabilidad de ganar una tirada.

d) En un número grande de jugadas, la proporción de veces que una persona gana tiene distribución normal *aproximada* con media = 0.493 y desviación estándar = $0.5\ \sqrt{n}$. Utilice esta información para calcular el número de jugadas simuladas que se requieren para obtener al menos una probabilidad de 0.95 de que la proporción de veces que gana sea menor a 0.5.

R **20.4-7.** Obtenga los números aleatorios uniformes como se indicó al principio, y utilice el método de transformación inversa y la tabla de la distribución normal del apéndice 5 (con interpolación lineal entre los valores de la tabla) para generar 10 observaciones aleatorias (con tres lugares decimales) a partir de una distribución normal con media = 1 y varianza = 4. Después calcule la media muestral de estas observaciones aleatorias.

R **20.4-8.** Obtenga números aleatorios uniformes del modo indicado al principio de los problemas, genere tres observaciones aleatorias (aproximadas) a partir de una distribución normal con media = 5 y desviación estándar = 10.

a) Aplique el teorema del límite central, con tres números aleatorios uniformes para generar cada observación.

b) Ahora utilice la tabla de la distribución normal del apéndice 5 y aplique el método de transformación inversa.

R **20.4-9.** Obtenga números aleatorios uniformes como se indicó al principio de los problemas, genere cuatro observaciones aleatorias (aproximadas) a partir de una distribución normal con media = 0 y desviación estándar = 1.

a) Aplique el teorema del límite central, con tres números aleatorios uniformes para generar cada observación.

b) Repita con la tabla de la distribución normal del apéndice 5 y aplique el método de transformación inversa.

c) Utilice las observaciones aleatorias en los incisos *a*) y *b*) para generar observaciones aleatorias a partir de una distribución ji cuadrada con 2 grados de libertad.

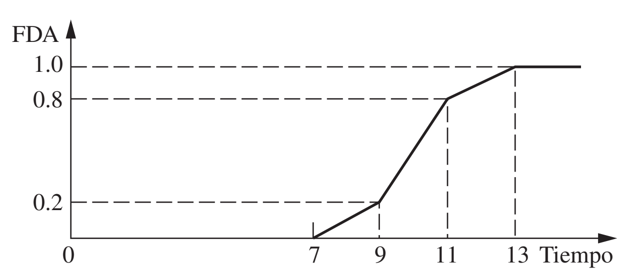
R **20.4-10.** Obtenga números aleatorios uniformes según se indicó, y genere dos observaciones aleatorias a partir de las siguientes distribuciones de probabilidad.

a) La distribución exponencial con media = 10.

b) La distribución Erlang con media = 10 y parámetro de forma $k = 2$ (esto es, desviación estándar $= 2\sqrt{2}$).

c) La distribución normal con media = 10 y desviación estándar $= 2\sqrt{2}$. (Use el teorema del límite central y $n = 6$ para cada observación.)

20.4-11. Richard Collins, director y dueño de Richard's Tire Service, desea usar simulación para analizar la operación de su tienda. Una actividad que debe incluir en la simulación es la instalación de llantas de automóvil (incluye el balanceo). Richard estima que la función de distribución acumulada (FDA) de la probabilidad del tiempo que se requiere (en minutos) para instalar una llanta tiene la siguiente gráfica.



a) Use la transformación inversa para generar 5 observaciones aleatorias a partir de esta distribución con los siguientes cinco

números aleatorios uniformes: 0.2655, 0.3472, 0.0248, 0.9205, 0.6130.

b) Use la función "SI" anidada para escribir una ecuación que use Excel para generar observaciones aleatorias a partir de esta distribución.

R **20.4-12.** Obtenga números aleatorios uniformes como se indicó al principio para generar cuatro observaciones aleatorias a partir de una distribución exponencial con media = 20. Después emplee estas observaciones para generar una observación aleatoria que siga una distribución Erlang con media = 4 y parámetro de forma $k = 4$.

20.4-13. Sean $r_1, r_2, \ldots, r_n$ números aleatorios uniformes. Defina $x_i = -\ln r_i$ y $y_i = -\ln(1 - r_i)$ para $i = 1, 2, \ldots, n$, y $z = \sum_{i=1}^{n} x_i$. Etiquete las siguientes afirmaciones como falsa o verdadera y justifique su respuesta.

a) Los números $x_1, x_2, \ldots, x_n$ y $y_1, y_2, \ldots, y_n$ son observaciones aleatorias de una distribución exponencial.

b) El promedio de $x_1, x_2, \ldots, x_n$ es igual al promedio de $y_1, y_2, \ldots, y_n$.

c) z es una observación aleatoria de una distribución Erlang (gamma).

20.4-14. Considere la variable aleatoria X que tiene una distribución uniforme (probabilidades iguales) en el conjunto $\{1, 2, \ldots, 8\}$. Se quiere generar una sucesión de observaciones aleatorias $x_i (i = 1, 2, \ldots)$ de X. Se han hecho las siguientes tres propuestas. Para cada una, analice si se trata de un método válido y, si no, diga cómo se puede ajustar para que lo sea.

a) Propuesta 1: generar números aleatorios uniformes r_i $(i = 1, 2, \ldots)$ y establecer $x_i = n$, donde n es un entero que satisface $n/8 \leq r_i < (n + 1)/8$.

b) Propuesta 2: generar números aleatorios uniformes r_i $(i = 1, 2, \ldots)$ y establecer x_i igual al entero mayor que es menor o igual a $1 + 8r_i$.

c) Propuesta 3: generar x_i congruencial mixto $x_{n+1} \equiv (5x_n + 7)$ (módulo 8), con un valor inicial $x_0 = 4$.

R **20.4-15.** Obtenga números aleatorios uniformes como se indicó al principio; use el método de aceptación-rechazo para generar tres observaciones aleatorias a partir de la distribución triangular que se usó en la sección 20.4.

R **20.4-16**. Obtenga números aleatorios uniformes como se indicó al principio de los problemas, y utilice el método de aceptación-rechazo para generar tres observaciones aleatorias a partir de la función de densidad de probabilidad

$$f(x) = \begin{cases} \dfrac{1}{50}(x - 10) & \text{si } 10 \leq x \leq 20 \\ 0 & \text{de otra manera.} \end{cases}$$

R **20.4-17.** Una compañía de seguros cubre grandes riesgos. El número de pérdidas en cada riesgo es independiente e idénticamente distribuido en los puntos $\{0, 1, 2\}$ con probabilidades respectivas de 0.7, 0.2 y 0.1. El tamaño de una pérdida individual tiene la siguiente función de distribución acumulada:

$$F(x) = \begin{cases} \dfrac{\sqrt{x}}{20} & \text{si } 0 \leq x \leq 100 \\ \dfrac{x}{200} & \text{si } 100 < x \leq 200 \\ 1 & \text{si } x > 200. \end{cases}$$

Obtenga números aleatorios según se indicó. Realice un experimento de simulación donde duplique el tamaño de la pérdida total generada por los cuatro riesgos.

20.4-18. Una compañía proporciona a sus 3 empleados un seguro de salud en un plan de grupo. Para cada empleado, la probabilidad de incurrir en gastos médicos durante el año es de 0.9, por lo cual el número de empleados que incurren en gastos médicos durante el año tiene distribución binomial con $p = 0.9$ y $n = 3$. Dado que un empleado incurre en gastos médicos durante un año, el monto total del año tiene la distribución de 100 dólares con probabilidad de 0.9, o 10 000 dólares con probabilidad de 0.1. La compañía tiene una cláusula de deducible de 5 000 dólares, de forma que cada año la aseguradora paga los gastos médicos totales del grupo que excedan esta cantidad. Utilice los números aleatorios 0.01 y 0.20, en ese orden, para generar el número de reclamaciones con base en una binomial para cada 2 años. Use los números aleatorios uniformes, en el orden dado, para generar el monto de cada reclamación: 0.80, 0.95, 0.70, 0.96, 0.54, 0.01. Calcule el monto total que paga la aseguradora en dos años.

20.5-1. Lea el artículo de referencia que describe el estudio de IO que se resume en el Recuadro de aplicación que se presentó en la sección 20.5. Describa de manera breve la forma en que se aplicó la simulación en este estudio. Después, elabore una lista de los beneficios financieros y de otro tipo que arrojó dicho estudio.

A **20.6-1** Los resultados de una corrida de simulación son inherentemente aleatorios. El presente problema demostrará este hecho e investigará el efecto del número de iteraciones sobre esta aleatoriedad. Considere el ejemplo que involucra a Freddie, el voceador, que se presentó en la sección 20.6. El modelo en hoja de cálculo se encuentra disponible en los archivos de Excel para este capítulo en el sitio en internet de este libro. Al usar Crystal Ball, tenga cuidado de no utilizar la misma secuencia de números aleatorios, esto es, *no* debe seleccionar la opción "Use Same Sequence of Random Number". Además, debe elegir el método de muestreo Monte-Carlo en la selección del muestreo del cuadro de diálogo "Run Preferences". Utilice un tamaño de orden de 60.

a) Establezca el número de iteraciones en 100 en el cuadro "Run Preferences" y corra la simulación del problema de Freddie cinco veces. Observe la ganancia media de cada corrida de simulación.

b) Repita el inciso *a*), pero ahora establezca el número de iteraciones en 1 000.

c) Compare los resultados de los incisos *a*) y *b*) y comente las diferencias.

A **20.6-2.** La Aberdeen Development Corporation (ADC) está reconsiderando el proyecto del Aberdeen Resort Hotel. Éste se localizaría en los pintorescos bancos de Grays Harbor y tendría su propio campo de golf de clase mundial.

El costo de la compra del terreno sería de 1 millón de dólares, que debe pagarse de inmediato. Los costos aproximados de la construcción serían de 2 millones de dólares, liquidables al final del año. Sin embargo, los costos de construcción son inciertos, es decir, podrían elevarse o reducirse hasta 20% a partir de la estimación de 2 millones de dólares. Suponga que los costos de construcción siguen una distribución triangular.

Existe mucha incertidumbre acerca de las ganancias (o pérdidas) de la operación anual que se generarían una vez que el hotel esté cons-

truido. Su mejor estimación de la ganancia por la operación anual que se generaría en los años 2, 3, 4 y 5 es de 700 000 dólares. Debido a la gran incertidumbre, la estimación de la desviación estándar de la ganancia anual en cada año también es de 700 000 dólares. Suponga que las ganancias anuales son estadísticamente independientes y siguen una distribución normal.

Después del año 5, ADC planea vender el hotel. Es probable que el precio de venta esté entre 4 y 8 millones de dólares (suponga una distribución uniforme). ADC utiliza una tasa de descuento de 10% para calcular el valor presente neto. (Para los propósitos de este cálculo, suponga que las ganancias anuales se reciben al final de cada año.) Realice mil iteraciones de una simulación de este proyecto en una hoja de cálculo.

a) ¿Cuál es la media del valor presente neto (VPN) del proyecto? [*Sugerencia:* La función VPN (tasa, valor) en Excel obtiene el VPN de una serie de flujos de efectivo que se supone que se iniciarán dentro de un año. Por ejemplo, VPN (10%, C5:F5) obtiene el valor presente neto a una tasa de descuento de 10% cuando C5 es un flujo de efectivo al final de año 1, D5 al final del año 2, E5 al final del año 3 y F5 al final del cuarto año].

b) ¿Cuál es la probabilidad estimada de que el proyecto produzca un VPN mayor a 2 millones de dólares?

c) La empresa también está preocupada por el flujo de efectivo en los años 2, 3, 4 y 5. Genere un pronóstico de la distribución de la ganancia anual *mínima* (antes de ser descontado) que se obtendrá en cualquiera de los cuatro años. ¿Cuál es el valor medio de la ganancia anual mínima a través de los cuatro años?

d) ¿Cuál es la probabilidad de que la ganancia anual sea al menos de 0 dólares en todos los años de operación?

A **20.6-3.** La fábrica Avery Co. ha tenido problemas de mantenimiento con el tablero de control de sus procesos de producción. El tablero contiene cuatro relevadores electromecánicos idénticos que han sido la causa del problema. Los relevadores fallan con frecuencia, lo cual provoca que el tablero se apague (y que el proceso de producción se detenga) mientras se hace el reemplazo. La práctica actual es reemplazar los relevadores sólo cuando fallan. El costo promedio de las sustituciones es de $3.19 por hora. Para intentar reducir este costo, se propone el reemplazo de los cuatro relevadores cada vez que uno falle para reducir la frecuencia con la que se apaga el tablero de control. ¿Realmente esta medida reducirá el costo?

Los datos pertinentes son: para cada relevador, el tiempo de operación antes de fallar tiene distribución uniforme aproximada de 1 000 a 2 000 horas. El tablero debe apagarse una hora para reemplazar un relevador y 2 para reemplazar los cuatro. El costo total asociado es de 1 000 dólares por hora más 200 dólares por cada nuevo relevador.

Use simulación en una hoja de cálculo para evaluar el costo de la propuesta y compárelo con el de la práctica actual. Realice 1 000 iteraciones (donde el final de cada una coincide con el final del reemplazo) y determine el costo promedio por hora.

A **20.6-4.** Para que la Aplus Company fabrique un nuevo producto es necesario perforar cojinetes en un bloque de metal e insertar ejes cilíndricos en los cojinetes. Se requiere que los ejes tengan al menos 1 pulgada de radio, y tan poco mayor que esto como sea posible. De acuerdo con el proceso de producción propuesto para los ejes, la distribución de probabilidad del radio tiene distribución triangular con un mínimo de 1 pulgada, un valor más probable de 1.0010 pulgadas y máximo de 1.0020 pulgadas. Según el proceso propuesto para perforar los cojinetes, la distribución de probabilidad del radio es normal con una media de 1.0020 pulgadas y una desviación estándar de 0.0010 pulgadas. El claro entre el cojinete y el eje es la diferencia de sus radios. Como se seleccionan al azar, suele haber obstrucción (claro negativo) entre el cojinete y el eje que se va a armar.

La administración está preocupada debido a que esta obstrucción en ocasiones provoca que la fabricación del nuevo producto se vea interrumpida. Quizá sea necesario mejorarse los procesos de producción de los ejes y cojinetes (a un costo alto) para reducir la posibilidad de obstrucción. Para evaluar la necesidad de esta mejora, le ha pedido que determine la frecuencia con la que ocurriría esta obstrucción con los procesos productivos que ahora se proponen.

Estime la probabilidad de obstrucción mediante 500 iteraciones de una simulación en hoja de cálculo.

A **20.6-5.** Reconsidere el problema 20.4-6 del juego de dados. El objetivo es estimar la probabilidad de ganar una jugada. Si la probabilidad es mayor de 0.5, usted irá a Las Vegas a jugar muchas veces hasta que gane una cantidad considerable de dinero; si es menor, se quedará en casa.

Decide realizar una simulación en hoja de cálculo para estimar esta probabilidad. Realice *dos veces* el número de iteraciones (jugadas) indicadas.

a) 100 iteraciones.

b) 1 000 iteraciones.

c) 10 000 iteraciones.

d) La probabilidad verdadera es de 0.493. ¿Qué conclusión obtiene de las corridas de simulación acerca del número de iteraciones que parece ser necesario para dar una seguridad razonable de obtener una estimación dentro de 0.007 de la probabilidad verdadera?

A **20.6-6.** Considere el ejemplo de Freddie, el voceador, que se introdujo en la sección 20.6. El modelo en hoja de cálculo está disponible en los archivos de Excel para este capítulo en el sitio en internet de este libro. La tabla de decisión que se generó en la sección 20.6 (vea la figura 20.19) para manejar el problema de Freddie sugiere que la mejor cantidad por ordenar es 55, pero esta tabla consideró sólo cantidades que son múltiplos de 5. Perfeccione la búsqueda y genere una tabla de decisión para el problema de Freddie que considere todas las cantidades por ordenar entre 50 y 60.

20.7-1. De la parte inferior de las referencias seleccionadas que se proporcionan al final del capítulo, seleccione una las aplicaciones de simulación ganadoras de premios. Lea ese artículo y después escriba un resumen de dos páginas de la aplicación y los beneficios (financieros y de otro tipo) que ella proporciona.

20.7-2. De la parte inferior de las referencias seleccionadas que se proporcionan al final del capítulo, seleccione tres de las aplicaciones de simulación ganadoras de premios. Para cada una de ellas, lea el artículo y después escriba un resumen de una página de la aplicación y los beneficios (financieros y de otro tipo) que ella proporciona.

CASOS

CASO 20.1 Reducción del inventario en proceso (modificado)

Reconsidere el caso 17.1. Los sistemas de colas actual y propuesto en este caso eran analizados con la ayuda de modelos de colas para determinar cómo reducir el inventario en proceso tanto como fuera posible. Sin embargo, estos mismos sistemas de colas también se pueden analizar de manera eficaz mediante simulación con la ayuda del simulador de colas del OR Courseware.

Use simulación para realizar todos los análisis que se solicitan en este caso.

CASO 20.2 Aventuras de acción

La compañía Adventure Toys fabrica una línea popular de figuras de acción y las distribuye en las tiendas de juguetes a un precio de mayoreo de 10 dólares cada una. La demanda es estacional, pues las ventas grandes ocurren antes de Navidad y durante la primavera. Las ventas más bajas ocurren durante los meses de verano e invierno (después de Navidad).

Cada mes, las ventas "bases" mensuales siguen una distribución normal con una media igual a la de las ventas "bases" reales del mes previo y una desviación estándar de 500 unidades. Las ventas reales en cualquier mes son las ventas bases mensuales multiplicadas por el factor estacional del mes, como se muestra en la tabla siguiente. Las ventas bases en diciembre de 2009 fueron de 6 000 con ventas reales igual a (1.18)(6 000) = 7 080. Ahora es el primero de enero de 2010.

Mes	Factor estacional	Mes	Factor estacional
Enero	0.79	Julio	0.74
Febrero	0.88	Agosto	0.98
Marzo	0.95	Septiembre	1.06
Abril	1.05	Octubre	1.10
Mayo	1.09	Noviembre	1.16
Junio	0.84	Diciembre	1.18

Las ventas en efectivo representan alrededor de 40% de las ventas mensuales, pero esta cifra ha descendido hasta 29% y se ha elevado hasta 48% en algunos meses. El resto de las ventas se hace a 30 días sin intereses, con un pago total que se debe recibir un mes después de la entrega. En diciembre de 2009, 42% de las ventas fueron en efectivo y 58% a crédito.

Los costos de producción dependen de la mano de obra y de los materiales. El precio de los plásticos necesarios para fabricar las figuras de acción fluctúan mes a mes, lo cual depende de las condiciones de mercado. Debido a estas fluctuaciones, los costos de producción pueden estar en cualquier punto entre 6 y 8 dólares por unidad. Además de estos costos de producción variables, la compañía incurre en un costo fijo de 15 000 dólares al mes por fabricar las figuras de acción. La compañía ensambla los productos por ordenar. Cuando se ordena un lote de una figura de acción particular, de inmediato se fabrica y embarca en un lapso no mayor a dos días.

La compañía utiliza ocho máquinas para moldear las figuras de acción. De manera ocasional estas máquinas se descomponen y requieren una parte de reemplazo que cuesta 5 000 dólares. Cada máquina necesita una parte de reemplazo cada mes con una probabilidad de 10%.

La compañía tiene como política mantener un saldo mínimo en efectivo de 20 000 dólares al final de cada mes. El saldo al final de diciembre de 2009 —o, de manera equivalente, a inicios de enero de 2010— es de 25 000 dólares. Si se necesita, la compañía puede solicitar un préstamo a corto plazo (1 mes) para cubrir los gastos y mantener el saldo mínimo. Los préstamos se deben pagar el mes siguiente con interés (se usa la tasa de interés para el préstamo del mes actual). Por ejemplo, si la tasa de interés anual de marzo es de 6% (Por lo tanto, 0.5% mensual) y se ha pedido un préstamo de 1 000 dólares en marzo, entonces en abril se deben 1 005 dólares. Sin embargo, se puede pedir un préstamo cada mes.

Cualquier saldo remanente al final de un mes (incluido el saldo mínimo) se conserva para el siguiente mes, y también genera intereses por ahorro. Por ejemplo, si el saldo final de marzo es de 20 000 dólares, y la tasa de interés de ahorro de ese mes es de 3% anual (es decir, 0.25% mensual), entonces en abril se gana 50 dólares de interés.

Tanto la tasa de interés de los préstamos como la tasa de ahorro se establece cada mes con base en la tasa Preferente (Prime rate). La tasa de interés de los préstamos es Preferente + 2%, mientras que la tasa de ahorro es Preferente – 2%. Sin embargo, la tasa de interés del préstamo está limitada a (no puede exceder a) 9% y la tasa de interés de ahorro no puede caer por debajo de 2%.

En diciembre de 2009 la tasa Preferente fue de 5% anual. Esta tasa depende de los caprichos de la Oficina de la Reserva Federal. En particular, para cada mes existe 70% de posibilidades de que ésta permanezca sin cambio, 10% de probabilidad de que aumente 0.25 puntos porcentuales, 10% de que disminuya 0.25 porcentuales, 5% de oportunidad de crecer 0.50 puntos porcentuales, y 5% de probabilidades de disminuir 0.50 puntos porcentuales.

a) Formule un modelo de simulación en hoja de cálculo para rastrear los flujos de efectivo mes a mes. Indique las distribuciones de probabilidad (tipo y parámetros) de las celdas de supuesto directamente en la hoja de cálculo. Simule 1 000 iteraciones para el año 2010, y pegue sus resultados en la hoja de cálculo.

b) La administración de Toys Adventure quiere información acerca de cuál será el valor neto de la compañía al final de 2010, lo que incluye la posibilidad de que el valor neto sea mayor que cero. (El valor neto se define aquí como el saldo de efectivo *más* los intereses de ahorro y cuentas por cobrar *menos* todos los préstamos e intereses por pagar.) Despliegue los resultados de su corrida de simulación del inciso *a*) en las diferentes formas que considere útiles para que la administración analice este asunto.

c) Se deben hacer arreglos para obtener un límite específico de crédito del banco para los créditos a corto plazo que podrían necesitarse durante 2010. Por lo tanto, la administración de Adventure Toys también desearía información acerca del tamaño del préstamo a corto plazo más grande que podría necesitar durante 2010. Despliegue los resultados de su corrida de simulación del inciso *a*) en las diferentes formas que considere útiles para que la administración analice este asunto.

RESUMEN DE LOS CASOS ADICIONALES EN EL SITIO EN INTERNET DE ESTE LIBRO (www.mhhe.com/hillier)

CASO 20.3 Planeación de aplanadores

El departamento de aplanados de una fábrica ha tenido dificultades para cumplir con su carga de trabajo, lo que ha afectado de manera seria el programa de producción de las siguientes operaciones. A veces el trabajo se acumula y se produce un gran atraso. Después puede haber largas pausas cuando no llega mucho trabajo, por lo que los aplanadores están ocupados sólo parte del tiempo. Se han presentado tres propuestas diferentes para evitar los cuellos de botella en el departamento de aplanados: 1) obtener un aplanador adicional, 2) eliminar la variabilidad del tiempo entre llegadas de los trabajos y 3) reducir la variabilidad del tiempo requerido para realizar los trabajos. Se puede adoptar cualquiera de estas propuestas o una combinación de ellas. Con la ayuda del simulador de colas, se usará simulación para determinar qué se debe hacer para minimizar el costo total esperado por hora.

CASO 20.4 Determinación de precios bajo presión

Un cliente de un gran banco de inversión está interesado en comprar una opción europea de cierta acción que le proporciona el derecho de comprar la acción a un precio fijo dentro de 12 semanas. Por lo tanto, el cliente ejercería esta opción en 12 semanas sólo si este precio fijo es menor que el precio de mercado de la acción en ese momento. Por su parte, el banco necesita determinar qué precio se debe cargar a la opción. Este precio debe ser el valor medio de la opción en 12 semanas. Con base en un modelo de "caminata" aleatoria que indica cómo evoluciona el precio de una acción semana a semana, se utilizará simulación para estimar este valor medio. Para iniciar, se deben formular con cuidado los diferentes elementos de un modelo de simulación.

APÉNDICE 1

Documentación del OR Courseware

Usted encontrará una gran variedad de recursos de software en el sitio web de este libro (www.mhhe.com/hillier). El paquete completo se llama *OR Courseware*.

Los paquetes de software se explican brevemente más adelante.

OR TUTOR

El OR Tutor es un documento web que consiste en una serie de páginas de HTML que con frecuencia contienen JavaScript. Se puede usar cualquier navegador con soporte de JavaScript, como Netscape Navigator 4.0 (o mayor) o Internet Explorer 4.5 (o mayor). También se puede ver en PC o en Macintosh.

Este recurso se diseñó como un tutor personal, que ilustra y resalta los conceptos claves de forma interactiva. Contiene 16 *ejemplos de demostración* que complementan los ejemplos del libro de una forma que no se logra en la página impresa. Cada uno muestra un algoritmo o concepto de IO en acción. La mayoría combina una *descripción algebraica* de cada paso con un *despliegue geométrico* de lo que ocurre. Algunas de las gráficas geométricas son dinámicas, con puntos o rectas que se mueven, para mostrar la evolución del algoritmo. Los ejemplos de demostración también están integrados con el libro, puesto que usan la misma notación y terminología, y se refieren al material del libro. Los estudiantes encuentran que es una ayuda amigable y eficaz para el aprendizaje.

IOR TUTORIAL

Otra característica importante del OR Courseware es un paquete de software llamado *Interactive Operations Research Tutorial* o *IOR Tutorial*. Éste es un producto de Accelet Corporation que se ha diseñado en forma específica para usarse con este libro. Se emplean características tutoriales innovadoras para que el proceso de aprendizaje de los algoritmos de este libro sea eficiente y amigable en la medida de lo posible. Está implementado en Java 2, por lo que se puede operar sobre cualquier plataforma. Vea las pantallas introductorias para conocer las instrucciones específicas de instalación.

El IOR Tutorial incluye una gran cantidad de *rutinas interactivas* para las diferentes áreas de conocimiento que cubre el libro. Cada una de estas rutinas permite *ejecutar* de manera *interactiva* uno de los algoritmos de IO. Mientras se ve la información relevante en la pantalla, puede tomar la decisión de cómo realizar el siguiente paso del algoritmo y después la computadora hace todos los cálculos necesarios para cumplir ese objetivo. Cuando se descubre un error cometido antes, la rutina le permite regresar para corregirlo. Para que inicie bien, la computadora señala cualquier error que tenga en la primera iteración (cuando es posible). Al terminar, puede imprimir todo el trabajo que realizó para entregarlo como tarea.

A juicio de los autores, estas rutinas interactivas proporcionan la forma "correcta" en esta era de las computadoras para que los estudiantes realicen sus tareas diseñadas para ayudarlo a comprender los algoritmos de IO. Las rutinas facilitan que se concentre en los conceptos y no en un sinfín de números sin sentido, para hacer que el proceso de aprendizaje sea más eficiente, eficaz y estimulante. También lo conducen en la dirección correcta, que incluye la organización de un trabajo bien hecho. Sin embargo, las rutinas no piensan por usted. Como en cualquier asignación de tarea bien diseñada, puede cometer errores (y aprender de ellos), de manera que necesitará pensar mucho para conservar la dirección correcta. Se ha tenido cuidado de crear una división del trabajo entre la computadora y el estudiante para que se pueda llevar a cabo un proceso de aprendizaje eficiente y completo.

Una vez que aprehenda la lógica de un algoritmo con la ayuda de una rutina interactiva, tendrá deseos de aplicar el algoritmo con rapidez de ahí en adelante, mediante una rutina automática. Este tipo de rutinas se encuentra en uno o más paquetes de software, que se analizarán más adelante, para la mayoría de los algoritmos descritos en el libro. No obstante, en el caso de ciertos algoritmos que no están incluidos en los paquetes comerciales (y para algunos que sí lo están), en el IOR Tutorial se proporcionan algunas rutinas automáticas especiales. Estas rutinas están diseñadas sólo para resolver los problemas de tamaño moderado incluidos en el libro.

ARCHIVOS DE EXCEL

El OR Courseware incluye archivos de Excel por separado para casi cada capítulo del libro. Lo normal es que el archivo incluya

varias hojas de cálculo que ayudan a formular y resolver los modelos descritos en el capítulo. Se incluyen dos tipos de hojas de cálculo. Primero, cada vez que se presenta un ejemplo que se puede resolver con Excel, se da la formulación completa y la solución. Esta ayuda proporciona una referencia conveniente, e incluso una plantilla útil, cuando desee resolver problemas similares con Excel Solver (o Premium Solver, que se verá en seguida). Segundo, para muchos modelos se proporcionan archivos con plantillas que ya incluyen las ecuaciones necesarias para resolverlos. Sólo tiene que introducir los datos y la solución se calcula de inmediato.

COMPLEMENTOS DE EXCEL

En el OR Courseware se incluyen cuatro complementos de Excel. Uno es el *Premium Solver for Education*, que es una versión más poderosa que el Solver estándar de Excel. Premium Solver es un producto de la misma organización que desarrolló Solver de Excel y también agregó Evolutionary Solver que se explica en la sección 12.10. Vea el sitio web del libro donde encontrará instrucciones para bajarlas del sitio www.solver.com de su desarrollador (Frontline Systems Inc.), pues ambas utilizan el código de texto (HLITOR) y el código de curso que necesita obtener el instructor (siga las instrucciones en el sitio web).

Tres de los otros complementos de Excel son las versiones académicas de *SensIt* (que se introdujo en la sección 15.5), *Tree-Plan* (sección 15.5) y *RiskSim* (sección 20.6). Todos son de uso común y fueron desarrollados por el profesor Michael R. Middleton para Windows y Macintosh. La documentación de los tres complementos se incluye en el sitio web del libro. Como el software es "shareware", quien desee usarlo después del curso debe registrarse y pagar la cuota.

Igual que con cualquier complemento de Excel, cada uno de ellos deben instalarse antes de operar. (Lo mismo ocurre con Excel Solver.) Las instrucciones de instalación se incluyen en el OR Courseware.

MPL/CPLEX

Como se estudió en las secciones 3.7 y 4.8, MPL es un lenguaje de modelado reciente y su solucionador CPLEX es en particular poderoso. Cualquier otro solucionador poderoso (que se describe en el siguiente párrafo) es compatible con MPL. La versión para estudiantes de MPL y CPLEX se incluyen en el OR Courseware. Aunque esta versión está limitada a problemas *mucho* más pequeños que los grandes problemas lineales, enteros y cuadráticos que resuelve la versión completa en la práctica, todavía puede manejar problemas *mucho* más grandes que los que se pueden encontrar en este libro.

El sitio web del libro proporciona un tutorial de MPL y documentación, así como formulaciones MPL/CPLEX y soluciones para casi todos los ejemplos a los que se aplica. Además, se encuentra la versión de estudiantes de OptiMax 2000, que permite integrar por completo los modelos de MPL en Excel y resolverlos con CPLEX. También, el solucionador de programación convexa CONOPT, el optimizador global LGO, el solucionador de programación lineal y entera CoinMP; el solucionador de programación lineal, entera y cuadrática LINDO; y el solucionador estocástico BendX se incluyen en MPL para resolver ese tipo de problemas.

La página de internet para conocer más de MPL y sus solucionadores o para bajar actualizaciones de MPL/CPLEX es www.maximalsoftware.com.

ARCHIVOS LINGO/LINDO

Este libro utiliza el popular lenguaje de modelado LINGO (vea especialmente el final de la Sección 3.7, los suplementos del capítulo 3 y el apéndice 4.1) incluyendo el tradicional subconjunto de sintaxis LINDO (vea sección 4.8 y apéndice 4.1). Versiones estudiantiles actualizadas de LINGO y LINDO (y el solucionador para hojas de cálculo *What's Best*) se pueden bajar de la página www.lindo.com.

El OR Courseware incluye un gran número de archivos de LINGO/LINDO o de LINGO (cuando LINDO no es relevante) para muchos capítulos. Cada archivo proporciona los modelos de LINDO y LINGO y las soluciones de los diferentes ejemplos del capítulo a los que se aplica. El sitio web del libro también contiene tutoriales de LINGO y LINDO.

ACTUALIZACIONES

El mundo del software evoluciona con mucha rapidez durante la vida de una edición de un libro de texto. Los autores consideran que la documentación que se proporciona en este apéndice es correcta en el momento de escribirla, pero sin duda ocurrirán cambios con el tiempo.

Para obtener información acerca de las actualizaciones de software, puede visitar la página del libro, www.mhhe.com/hillier.

APÉNDICE 2

Convexidad

Como se explicó en el capítulo 12, el concepto de *convexidad* se usa en el trabajo de IO, en especial en el área de programación no lineal. Por lo tanto, a continuación se presentan las propiedades de las funciones convexas o cóncavas y de los conjuntos convexos.

FUNCIONES CONVEXAS O CÓNCAVAS DE UNA SOLA VARIABLE

Para comenzar se presentan las definiciones.

> **Definición:** una *función* de una sola variable, $f(x)$, es una **función convexa** si, para *cada* par de valores de x, por ejemplo, x' y x'' ($x' < x''$),
>
> $$f[\lambda x'' + (1 - \lambda)x'] \leq \lambda f(x'') + (1 - \lambda)f(x')$$
>
> para todos los valores de λ tales que $0 < \lambda < 1$. Ésta es una **función estrictamente convexa** si $\leq$ puede sustituirse por $<$. Es una **función cóncava** (o una **función estrictamente cóncava**) si esta afirmación se cumple cuando se reemplaza $\leq$ por $\geq$ (o por $>$).

La definición de función convexa tiene una interpretación geométrica muy ilustrativa. Considere la gráfica de la función $f(x)$ que se trazó como función de x en la figura A2.1 para una función $f(x)$ que decrece para $x < 1$, es constante para $1 \leq x \leq 2$ y crece para $x > 2$. Entonces $[x', f(x')]$ y $[x'', f(x'')]$ son dos puntos de la gráfica de $f(x)$ y $[\lambda x'' + (1 - \lambda)x', \lambda f(x'') + (1 - \lambda) f(x')]$ representa los diferentes puntos sobre el segmento rectilíneo entre estos dos puntos (pero excluye los puntos terminales), cuando $0 < \lambda < 1$. Así, la desigualdad del tipo $\leq$ en la definición indica que este segmento rectilíneo está todo por encima o sobre la gráfica de la función, como se ve en la figura A2.1. En consecuencia, $f(x)$ es *convexa* si, para *cada* par de puntos sobre la gráfica de $f(x)$ el segmento de línea que los une se encuentra completamente por encima o sobre la gráfica de $f(x)$.

Por ejemplo, el resultado de la elección específica de x' y x'' de la figura A2.1 es un segmento de recta que (excepto por los puntos terminales) está todo *por encima* de la gráfica de $f(x)$. Esto ocurre también en el caso de otros pares de puntos x' y x'' donde $x' < 1$ o $x'' > 2$ (o ambas condiciones). Si $1 \leq x' < x'' \leq 2$, todo el segmento de recta está *sobre* la gráfica de $f(x)$. Por lo tanto, $f(x)$ es convexa.

Esta interpretación geométrica indica que $f(x)$ es convexa si y sólo si se "dobla hacia arriba", si es que se dobla. (En ocasiones, esta condición se conoce como *cóncava hacia arriba*, al contrario de *cóncava hacia abajo* de una función cóncava.) Para ser más precisos, si $f(x)$ tiene una segunda derivada en cualquier punto, entonces $f(x)$ es cóncava si y sólo si $d^2f(x)/dx^2 \geq 0$ para todos los valores posibles de x.

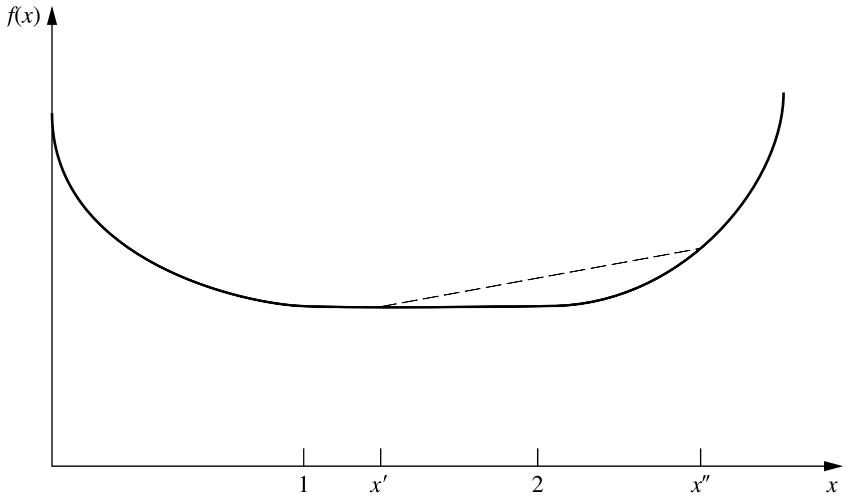
Las definiciones de una *función estrictamente convexa*, una *función cóncava* y una *función estrictamente cóncava* tienen interpretaciones geométricas análogas que se resumen en términos de la segunda derivada de la función, y es una prueba conveniente del estado de la función.

> **Prueba de convexidad de una función de una variable:** Considere cualquier función de una variable $f(x)$ que tiene una segunda derivada para todos los valores posibles de x. Entonces $f(x)$ es
>
> 1. *Convexa* si y sólo si $\dfrac{d^2f(x)}{dx^2} \geq 0$ para todos los valores posibles de x
> 2. *Estrictamente convexa* si y sólo si $\dfrac{d^2f(x)}{dx^2} > 0$ para todos los valores posibles de x
> 3. *Cóncava* si y sólo si $\dfrac{d^2f(x)}{dx^2} \leq 0$ para todos los valores posibles de x
> 4. *Estrictamente cóncava* si y sólo si $\dfrac{d^2f(x)}{dx^2} < 0$ para todos los valores posibles de x

Observe que una función estrictamente convexa también es convexa, pero una función convexa *no* es estrictamente convexa si la segunda derivada es igual a cero para algunos valores de x. De igual manera, una función estrictamente cóncava es cóncava, pero el inverso puede no ser cierto.

Las figuras A2.1 a A2.6 dan ejemplos que ilustran estas definiciones y la prueba de convexidad.

Al aplicar esta prueba a la función de la figura A2.1, se ve que cuando x crece, la pendiente (primera derivada) crece (para



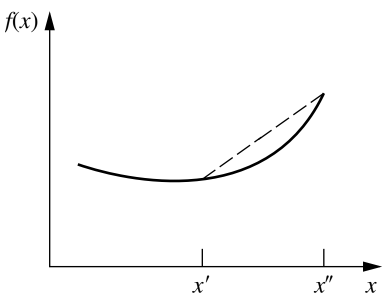
■ **FIGURA A2.1**
Función convexa.

$0 \le x < 1$ y $x > 2$) o permanece constante (para $1 \le x_1 \le 2$). Por lo tanto, la segunda derivada siempre es no negativa, lo que verifica que la función es convexa. Sin embargo, *no* es estrictamente convexa porque la segunda derivada es igual a cero para $1 \le x \le 2$.

De cualquier manera, la función de la figura A2.2 es estrictamente convexa porque su pendiente siempre aumenta, con lo que la segunda derivada siempre es mayor que cero.

La función seccionalmente lineal que se muestra en la figura A2.3 cambia su pendiente en $x = 1$. En consecuencia, no tiene primera ni segunda derivadas en este punto, y la prueba de convexidad no se puede aplicar en todos los puntos. (El hecho de que la segunda derivada sea igual a cero para $0 \le x < 1$ y $x > 1$ hace que la función sea elegible como convexa o bien como cóncava, según su comportamiento en $x = 1$.) Al aplicar la definición de función cóncava se ve que si $0 < x' < 1$ y $x'' > 1$ (como se muestra en la figura A2.3), entonces todo el segmento de recta que une a $[x', f(x')]$ y $[x'', f(x'')]$ está *abajo* de la gráfica de $f(x)$, excepto por los puntos terminales del segmento. Si ocurre cualquiera de $0 \le x' < x'' \le 1$ o $1 \le x' < x''$, entonces el segmento completo está *sobre* la gráfica de $f(x)$ y, por lo tanto, $f(x)$ es cóncava (pero *no* estrictamente cóncava).

■ **FIGURA A2.2**
Función estrictamente convexa.



La función de la figura A2.4 es estrictamente cóncava porque su segunda derivada siempre es menor que cero.

Como se ilustra en la figura A2.5, cualquier función lineal tiene su segunda derivada igual a cero en cualquier punto, por lo cual es cóncava y convexa.

La función de la figura A2.6 no es cóncava *ni* convexa porque conforme x crece, la pendiente fluctúa entre crecer y decrecer por lo que la segunda derivada se maneja entre valores positivos y negativos.

FUNCIONES CONVEXAS O CÓNCAVAS DE VARIAS VARIABLES

El concepto de función convexa o cóncava de una sola variable se generaliza también a funciones de más de una variable. De este modo, si se sustituye $f(x)$ por $f(x_1, x_2, \ldots, x_n)$, la definición que se acaba de dar se aplica al sustituir x en todas partes por $(x_1, x_2, \ldots, x_n)$. De manera similar, las interpretaciones geométricas correspondientes son válidas después de generalizar los conceptos de *puntos* y *segmentos rectilíneos*. Así, igual que un valor particular de (x, y) se interpreta como un punto en el espacio bidimensional, cada valor posible de $(x_1, x_2, \ldots, x_m)$ puede concebirse como un punto en el espacio (euclidiano) m-dimensional. Al hacer $m = n + 1$, los puntos sobre la gráfica de $f(x_1, x_2, \ldots, x_n)$ se convierten en los valores posibles de $[x_1, x_2, \ldots, x_n, f(x_1, x_2, \ldots, x_n)]$. Se dice que otro punto $(x_1, x_2, \ldots, x_n, x_{n+1})$ está arriba, sobre o abajo de la gráfica de $f(x_1, x_2, \ldots, x_n)$, según si el valor de x_{n+1} es mayor, igual o menor que $f(x_1, x_2, \ldots, x_n)$, respectivamente.

Definición: El **segmento de recta** que une cualesquiera dos puntos $(x'_1, x'_2, \ldots, x'_m)$ y $(x''_1, x''_2, \ldots, x''_m)$ es la colección de puntos

$$(x_1, x_2, \ldots, x_m) = [\lambda x''_1 + (1-\lambda)x'_1, \lambda x''_2 + (1-\lambda)x'_2, \ldots, \lambda x''_m + (1-\lambda)x'_m]$$

tales que $0 \le \lambda \le 1$.

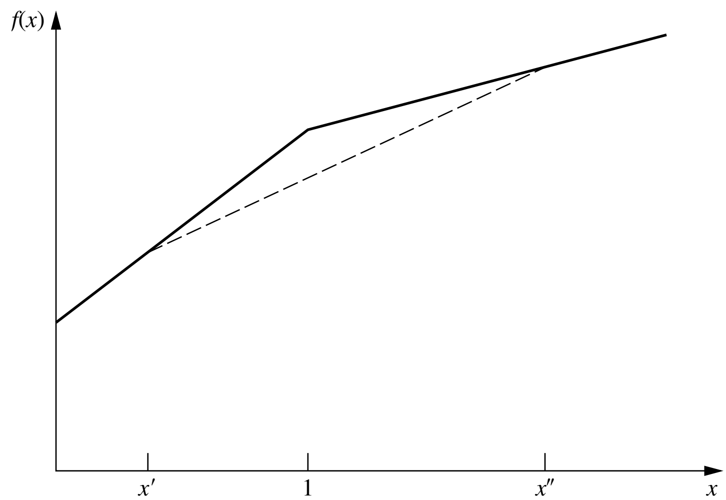


FIGURA A2.3
Función cóncava.

Así, un segmento de recta en el espacio *m*-dimensional es una generalización directa de un segmento de recta en el espacio de dos dimensiones. Por ejemplo, si

$$(x_1', x_2') = (2, 6), \qquad (x_1'', x_2'') = (3, 4),$$

por lo que el segmento de recta que los une es la colección de puntos

$$(x_1, x_2) = [3\lambda + 2(1 - \lambda), 4\lambda + 6(1 - \lambda)],$$

donde $0 \leq \lambda \leq 1$.

Definición: $f(x_1, x_2, \ldots, x_n)$ es una **función convexa** si, para cada par de puntos sobre la gráfica de $f(x_1, x_2, \ldots, x_n)$, el segmento de recta que los une se encuentra completamente arriba o sobre la gráfica de $f(x_1, x_2, \ldots, x_n)$. Es una **función estrictamente convexa** si este segmento de recta está por completo arriba de la gráfica, excepto por los puntos extremos del segmento. Las **funciones cóncavas** y las **funciones estrictamente cóncavas** se definen exactamente de la misma manera, excepto que la palabra *arriba* debe sustituirse por *abajo*.

De la misma manera en que puede usarse la segunda derivada (cuando existe en todas partes) para comprobar si una función de una sola variable es convexa o no, las segundas derivadas parciales pueden emplearse para ver qué pasa con funciones de varias variables, aunque en una forma más complicada. Por ejemplo, si se tienen dos variables y todas las derivadas parciales existen en todas partes, entonces la prueba de convexidad evalúa si *las tres cantidades* dadas en la primera columna de la tabla A2.1 satisfacen las desigualdades que se muestran en la columna apropiada para *todos los valores posibles* de (x_1, x_2).

Cuando se tiene más de dos variables, la prueba de convexidad es una generalización de la que se presenta en la tabla A2.1. Por ejemplo, en terminología de matemáticas, $f(x_1, x_2, \ldots, x_n)$ es convexa si y sólo si su matriz Hessiana de $n \times n$ es positiva semidefinida para todos los valores posibles de $(x_1, x_2, \ldots, x_n)$.

FIGURA A2.4
Función estrictamente cóncava.

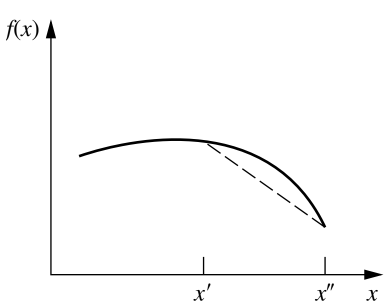


FIGURA A2.5
Función que es tanto cóncava como convexa.

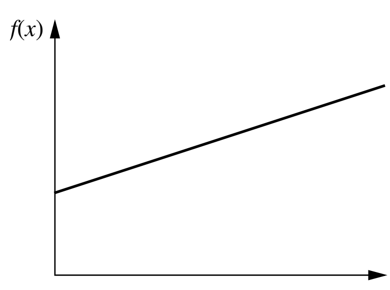
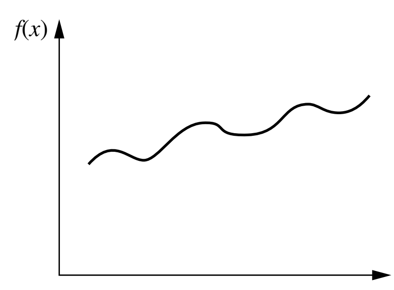


FIGURA A2.6
Función que no es cóncava ni convexa.



Para ilustrar la prueba de convexidad de dos variables, considere la función

$$f(x_1, x_2) = (x_1 - x_2)^2 = x_1^2 - 2x_1x_2 + x_2^2.$$

Por lo tanto,

$$(1) \quad \frac{\partial^2 f(x_1, x_2)}{\partial x_1^2} \frac{\partial^2 f(x_1, x_2)}{\partial x_2^2} - \left[\frac{\partial^2 f(x_1, x_2)}{\partial x_1 \partial x_2}\right]^2 = 2(2) - (-2)^2 = 0,$$

$$(2) \quad \frac{\partial^2 f(x_1, x_2)}{\partial x_1^2} = 2 > 0,$$

$$(3) \quad \frac{\partial^2 f(x_1, x_2)}{\partial x_2^2} = 2 > 0.$$

Como ≥ 0 se cumple para las tres condiciones, $f(x_1, x_2)$ es convexa. Sin embargo, *no* es estrictamente convexa porque la primera condición sólo da $= 0$ en lugar de > 0.

Ahora considere el negativo de esta función

$$g(x_1, x_2) = -f(x_1, x_2) = -(x_1 - x_2)^2 = -x_1^2 + 2x_1x_2 - x_2^2.$$

En este caso,

$$(4) \quad \frac{\partial^2 g(x_1, x_2)}{\partial x_1^2} \frac{\partial^2 g(x_1, x_2)}{\partial x_2^2} - \left[\frac{\partial^2 g(x_1, x_2)}{\partial x_1 \partial x_2}\right]^2 = -2(-2) - 2^2 = 0,$$

$$(5) \quad \frac{\partial^2 g(x_1, x_2)}{\partial x_1^2} = -2 < 0,$$

$$(6) \quad \frac{\partial^2 g(x_1, x_2)}{\partial x_2^2} = -2 < 0.$$

Como para la primera condición se cumple ≥ 0 y para las otras dos se cumple ≤ 0, $g(x_1, x_2)$ es una función cóncava. Sin embargo, *no* es estrictamente cóncava puesto que la primera condición da $= 0$.

Hasta aquí se ha tratado la convexidad como una propiedad general de una función. Sin embargo, muchas funciones no convexas también satisfacen las condiciones de convexidad sobre ciertos intervalos de las variables respectivas. Por lo tanto, tiene sentido hablar de una función convexa sobre cierta región. Por ejemplo, se dice que una función es convexa en la vecindad de un punto específico si su segunda derivada o sus derivadas parciales satisfacen las condiciones de convexidad en ese punto. Este concepto es de utilidad en el apéndice 3.

Por último, deben mencionarse dos propiedades en particular importantes de las funciones convexas. Primero, si $f(x_1, x_2, \ldots, x_n)$ es convexa, entonces $g(x_1, x_2, \ldots, x_n) = -f(x_1, x_2, \ldots, x_n)$ es cóncava, y viceversa, como se ilustra en el ejemplo anterior donde $f(x_1, x_2) = (x_1 - x_2)^2$. Segundo, la suma de funciones convexas es convexa y la suma de funciones cóncavas es cóncava. Como ilustración,

$$f_1(x_1) = x_1^4 + 2x_1^2 - 5x_1$$

y

$$f_2(x_1, x_2) = x_1^2 + 2x_1x_2 + x_2^2$$

son, ambas, funciones convexas, como se puede verificar si se calculan sus segundas derivadas. En consecuencia, la suma de estas funciones,

$$f(x_1, x_2) = x_1^4 + 3x_1^2 - 5x_1 + 2x_1x_2 + x_2^2$$

es una función convexa, mientras que su negativo,

$$g(x_1, x_2) = -x_1^4 - 3x_1^2 + 5x_1 - 2x_1x_2 - x_2^2,$$

es una función cóncava.

CONJUNTOS CONVEXOS

El concepto de una función convexa lleva de una manera bastante natural al concepto relacionado de **conjunto convexo**. Así, si $f(x_1, x_2, \ldots, x_n)$ es una función convexa, la colección de puntos que se encuentran arriba o sobre la gráfica de $f(x_1, x_2, \ldots, x_n)$ forman un conjunto convexo. De igual modo, la colección de puntos que se encuentran abajo o sobre la gráfica de una función cóncava es un conjunto convexo. Estos casos se ilustran en las figuras A2.7 y A2.8 cuando se tiene una sola variable independiente. Más aún, los conjuntos convexos tienen la importante propiedad de que, para cualquier grupo dado de conjuntos convexos, la colección de puntos que se encuentran en todos ellos (es decir, su intersección) también es un conjunto convexo. Por lo tanto, la colección de puntos que están tanto arriba como

TABLA A2.1 Prueba de convexidad de una función de dos variables.

Cantidad	Convexa	Estrictamente convexa	Cóncava	Estrictamente cóncava
$\frac{\partial^2 f(x_1, x_2)}{\partial x_1^2} \frac{\partial^2 f(x_1, x_2)}{\partial x_2^2} - \left[\frac{\partial^2 f(x_1, x_2)}{\partial x_1 \partial x_2}\right]^2$	≥ 0	> 0	≥ 0	> 0
$\frac{\partial^2 f(x_1, x_2)}{\partial x_1^2}$	≥ 0	> 0	≤ 0	< 0
$\frac{\partial^2 f(x_1, x_2)}{\partial x_2^2}$	≥ 0	> 0	≤ 0	< 0
Valores de (x_1, x_2)	Todos los valores posibles			

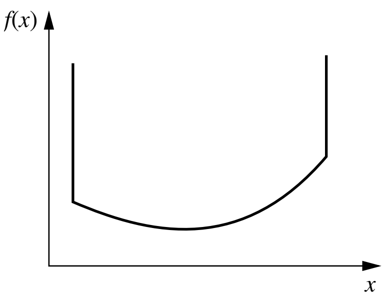


FIGURA A2.7
Ejemplo de un conjunto convexo determinado por una función convexa.

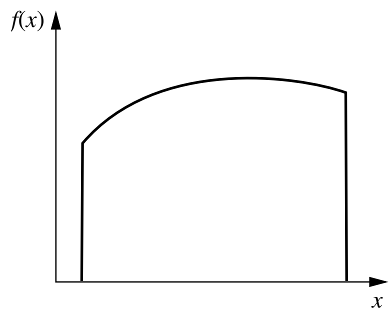


FIGURA A2.8
Ejemplo de un conjunto convexo determinado por una función cóncava.

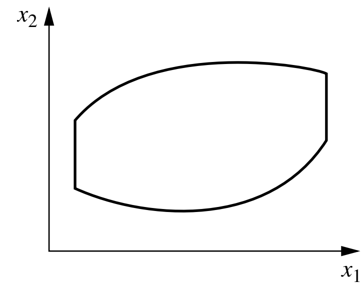


FIGURA A2.9
Ejemplo de un conjunto convexo determinado tanto por una función convexa como por una cóncava.

sobre una función convexa, y tanto abajo como sobre una función cóncava, es un conjunto convexo, como se ve en la figura A2.9. Por lo tanto, los conjuntos convexos pueden concebirse de manera intuitiva como una colección de puntos cuya frontera inferior es una función convexa y cuya frontera superior es una función cóncava.

Aunque puede ser útil describir los conjuntos convexos en términos de funciones cóncavas y convexas para desarrollar la intuición sobre su naturaleza, sus definiciones actuales no tienen nada que ver (directamente) con tales funciones.

Definición: Un **conjunto convexo** es una colección de puntos tales que, para cada par de puntos de la colección, el segmento rectilíneo completo que une estos dos puntos también está en la colección.

La diferencia entre conjuntos no convexos y conjuntos convexos se ilustra en las figuras A2.10 y A2.11. Así, el conjunto de puntos que se muestra en la figura A2.10 no es un conjunto convexo porque existen muchos pares de esos puntos, por ejemplo, (1, 2) y (2, 1), tales que el segmento rectilíneo entre ellos no está por completo dentro del conjunto. Éste no es el caso del conjunto de la figura A2.11, que es convexo.

En conclusión, se introduce el útil concepto de punto extremo de un conjunto convexo.

Definición: Un **punto extremo** de un conjunto convexo es un punto del conjunto que no está en ningún segmento rectilíneo que une a otros dos puntos del conjunto.

Entonces, los puntos extremos del conjunto convexo de la figura A2.11 son (0, 0), (0, 2), (1, 2), (2, 1), (1, 0) y el número infinito de puntos sobre la frontera entre (2, 1) y (1, 0). Si, por el contrario, esta frontera en particular fuera un segmento de recta, entonces el conjunto sólo tendría los cinco puntos extremos que se mencionaron.

FIGURA A2.10
Ejemplo de un conjunto no convexo.

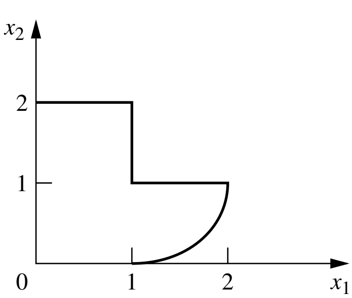
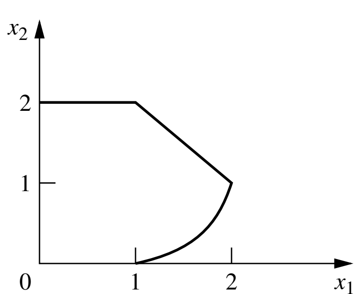


FIGURA A2.11
Ejemplo de un conjunto convexo.



APÉNDICE 3

Métodos de optimización clásica

En este apéndice se exponen en forma breve los métodos clásicos del cálculo para encontrar una solución que maximice o minimice: 1) una función de una sola variable, 2) una función de varias variables y 3) una función de varias variables sujeta a restricciones respecto de los valores de estas variables. Se supone que las funciones que se estudian tienen primera y segunda derivadas continuas y derivadas parciales en todas partes. Algunos de los conceptos que se presentan aquí se introdujeron en las secciones 12.2 y 12.3.

OPTIMIZACIÓN NO RESTRINGIDA DE UNA FUNCIÓN DE UNA SOLA VARIABLE

Considere una función de una variable como la que se muestra en la figura A3.1. Una condición necesaria para que una solución particular, $x = x^*$, sea un mínimo o un máximo es que

$$\frac{df(x)}{dx} = 0 \qquad \text{en } x = x^*.$$

Así, en la figura A3.1 existen cinco soluciones que satisfacen estas condiciones. Para obtener más información acerca de estos cinco **puntos críticos**, es necesario examinar la segunda derivada. Por ello, si

$$\frac{d^2f(x)}{dx^2} > 0 \qquad \text{en } x = x^*,$$

entonces x^* debe ser por lo menos un **mínimo local** [esto es, $f(x^*) \leq f(x)$ para toda x suficientemente cercana a x^*]. En el lenguaje que se introdujo en el apéndice 2, esto afirma que x^* debe ser un mínimo local si $f(x)$ es *estrictamente convexa*, en una vecindad de x^*. De la misma manera, una condición suficiente para que x^* sea un **máximo local** (dado que satisface la condición necesaria) es que $f(x)$ sea *estrictamente cóncava* en una vecindad de x^* (la segunda derivada es *negativa* en x^*). Si la segunda derivada es cero, no se llega a una solución (el punto incluso puede ser *un punto de inflexión*) y será necesario examinar derivadas superiores.

Para encontrar un **mínimo global** [o sea, una solución x^* tal que $f(x^*) \leq f(x)$ para toda x], es necesario comparar los mínimos locales e identificar el que proporcione el menor valor de $f(x)$. Si este valor es menor que $f(x)$ cuando $x \to -\infty$ y cuando $x \to +\infty$ (o en los puntos terminales de la función, si sólo está definida para un intervalo finito), entonces este punto es un mínimo global. En la figura A3.1 se muestra un punto de este tipo, junto con el **máximo global** identificado de manera análoga.

Sin embargo, si se sabe que $f(x)$ es una función convexa o cóncava (vea en el apéndice 2 una descripción de tales funciones), el análisis se vuelve mucho más sencillo. En particular, si $f(x)$ es una función *convexa*, como la que se muestra en la figura A2.1, entonces se sabe que cualquier solución x^*, tal que

$$\frac{df(x)}{dx} = 0 \qquad \text{en } x = x^*$$

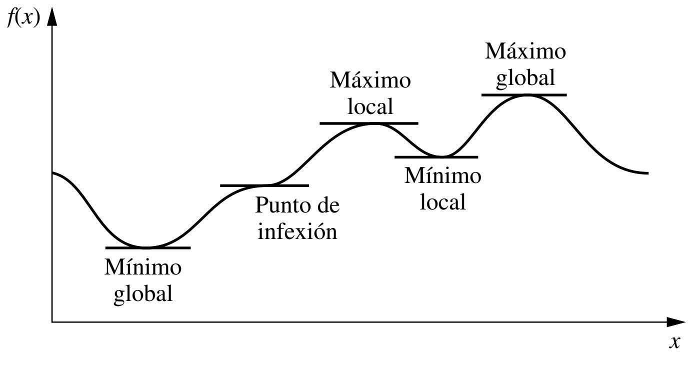


FIGURA A3.1
Función que tiene varios máximos y mínimos.

es un *mínimo global*. En otras palabras, esta condición no sólo es *necesaria* sino también *suficiente* para la existencia de un mínimo global de la función convexa. Esta solución no tiene que ser única, puesto que puede haber un empate para el mínimo global en un intervalo donde la derivada es cero. Por otro lado, si $f(x)$ en realidad es *estrictamente convexa*, entonces esta solución debe ser el único mínimo global. (No obstante, si la función es siempre creciente o siempre decreciente, de modo que la derivada es distinta de cero para todos los valores de x, entonces no existirá mínimo global para un valor finito de x.)

En forma similar, si $f(x)$ es una función *cóncava*, entonces

$$\frac{df(x)}{dx} = 0 \qquad \text{en } x = x^*$$

se convierte en una condición tanto *necesaria* como *suficiente* para que x^* sea un *máximo global*.

OPTIMIZACIÓN NO RESTRINGIDA DE UNA FUNCIÓN DE VARIAS VARIABLES

El análisis de una función no restringida de varias variables $f(\mathbf{x})$, donde $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ es parecido. Así, una condición *necesaria* para que una solución $\mathbf{x} = \mathbf{x}^*$ sea un mínimo o un máximo es que

$$\frac{\partial f(\mathbf{x})}{\partial x_j} = 0 \qquad \text{en } \mathbf{x} = \mathbf{x}^*, \text{ para } j = 1, 2, \ldots, n.$$

Después de identificar los puntos críticos que satisfacen esta condición, cada uno de ellos se clasifica como un mínimo o un máximo local si la función es *estrictamente convexa* o *estrictamente cóncava*, respectivamente, en la vecindad del punto. (Se requiere un análisis adicional si la función no es de alguno de estos dos tipos.) El *mínimo* y el *máximo globales* se encontrarían al comparar los mínimos y máximos relativos o locales y al verificar después el valor de la función cuando algunas de las variables tienden a $-\infty$ o $+\infty$. Sin embargo, si se sabe que la función es *convexa* o *cóncava*, entonces un punto crítico debe ser un *mínimo global* o un *máximo global*, respectivamente.

OPTIMIZACIÓN RESTRINGIDA CON RESTRICCIONES DE IGUALDAD

Ahora considere el problema de encontrar el *mínimo* o *máximo* de la función $f(\mathbf{x})$, sujeta a la restricción de que $\mathbf{x}$ debe satisfacer todas las ecuaciones

$$\begin{aligned} g_1(\mathbf{x}) &= b_1 \\ g_2(\mathbf{x}) &= b_2 \\ &\vdots \\ g_m(\mathbf{x}) &= b_m, \end{aligned}$$

donde $m < n$. Por ejemplo, si $n = 2$ y $m = 1$, el problema podría ser

$$\text{Maximizar} \qquad f(x_1, x_2) = x_1^2 + 2x_2,$$

sujeta a

$$g(x_1, x_2) = x_1^2 + x_2^2 = 1.$$

En este caso, (x_1, x_2) está restringida a encontrarse sobre el círculo de radio 1, cuyo centro está en el origen, de forma que la meta es encontrar el punto sobre este círculo que proporcione el mayor valor de $f(x_1, x_2)$. Es sencillo resolver este ejemplo después de bosquejar un enfoque general del problema.

Un método clásico para manejar este problema es el **método de los multiplicadores de Lagrange**. Este procedimiento comienza por plantear la **función Lagrangiana**

$$h(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i[g_i(\mathbf{x}) - b_i],$$

donde las nuevas variables $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ se llaman *multiplicadores de Lagrange*. Observe el hecho fundamental de que para los valores *factibles* de $\mathbf{x}$,

$$g_i(\mathbf{x}) - b_i = 0, \qquad \text{para toda } i,$$

entonces $h(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x})$. Por lo tanto, se puede demostrar que si $(\mathbf{x}, \boldsymbol{\lambda}) = (\mathbf{x}^*, \boldsymbol{\lambda}^*)$ es un *mínimo* o *máximo local* o *global* de la función no restringida $h(\mathbf{x}, \boldsymbol{\lambda})$, entonces $\mathbf{x}^*$ es el *punto crítico* correspondiente del problema original. Como resultado, el método se reduce ahora al análisis de $h(\mathbf{x}, \boldsymbol{\lambda})$ por el procedimiento descrito para optimización no restringida. Se igualan a cero las $n + m$ derivadas parciales

$$\frac{\partial h}{\partial x_j} = \frac{\partial f}{\partial x_j} - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} = 0, \qquad \text{para } j = 1, 2, \ldots, n,$$

$$\frac{\partial h}{\partial \lambda_i} = -g_i(\mathbf{x}) + b_i = 0, \qquad \text{para } i = 1, 2, \ldots, m,$$

y entonces se obtienen los puntos críticos al despejar $(\mathbf{x}, \boldsymbol{\lambda})$ de estas ecuaciones. Observe que las últimas m ecuaciones son equivalentes a las restricciones del problema original, de manera que sólo se consideran las soluciones factibles. Después del análisis adicional para identificar el *máximo* o *mínimo* global de $h(\cdot)$, el valor de $\mathbf{x}$ que resulta es la solución deseada del problema original.

Desde un punto de vista práctico y de cálculo, el método de los multiplicadores de Lagrange no es un procedimiento particularmente eficaz. Con frecuencia es en esencia imposible resolver las ecuaciones para obtener los puntos críticos. Es más, aun cuando sea posible obtenerlos, el número de puntos críticos puede ser tan grande (a menudo infinito) que no resulta práctico intentar la identificación de un máximo o un mínimo global. Sin embargo, para ciertos tipos de problemas pequeños, algunas veces puede usarse este método con éxito.

A manera de ilustración, considere el ejemplo que se presentó antes. En este caso

$$h(x_1, x_2) = x_1^2 + 2x_2 - \lambda(x_1^2 + x_2^2 - 1),$$

de manera que

$$\frac{\partial h}{\partial x_1} = 2x_1 - 2\lambda x_1 = 0,$$

$$\frac{\partial h}{\partial x_2} = 2 - 2\lambda x_2 = 0,$$

$$\frac{\partial h}{\partial \lambda} = -(x_1^2 + x_2^2 - 1) = 0.$$

La primera ecuación implica que $\lambda = 1$ o $x_1 = 0$. Si $\lambda = 1$, entonces las otras dos ecuaciones implican que $x_2 = 1$ y $x_1 = 0$. Si $x_1 = 0$, entonces la tercera ecuación implica que $x_2 = \pm 1$. Por lo tanto, los dos puntos críticos del problema original son $(x_1, x_2) = (0, 1)$ y $(0, -1)$. En consecuencia, es evidente que estos dos puntos son los respectivos máximo y mínimo globales.

DERIVADA DE UNA INTEGRAL DEFINIDA

Al presentar los métodos de optimización clásica que acaban de describirse, se ha supuesto que el lector está familiarizado con las derivadas y su obtención. Sin embargo, existe un caso especial importante en el trabajo de IO, que justifica una explicación adicional, a saber, la derivada de una integral definida. En particular, considere el problema de cómo encontrar la derivada de la función

$$F(y) = \int_{g(y)}^{h(y)} f(x, y)\, dx,$$

donde $g(y)$ y $h(y)$ son los límites de integración expresados como funciones de y.

Primero, suponga que estos límites de integración son constantes, de manera que $g(y) = a$ y $h(y) = b$, respectivamente. En este caso especial puede demostrarse que, dadas las condiciones de regularidad supuestas al principio de este apéndice, la derivada es sólo

$$\frac{d}{dy} \int_a^b f(x, y)\, dx = \int_a^b \frac{\partial f(x, y)}{\partial y}\, dx.$$

Por ejemplo, si $f(x, y) = e^{-xy}$, $a = 0$ y $b = \infty$, entonces,

$$\frac{d}{dy} \int_0^\infty e^{-xy}\, dx = \int_0^\infty (-x)e^{-xy}\, dx = -\frac{1}{y^2}$$

para cualquier valor positivo de y. De esta forma, el procedimiento intuitivo para intercambiar el orden de derivación e integración es válido en este caso.

Sin embargo, encontrar la derivada se vuelve un poco más complicado cuando los límites de integración son funciones. En particular,

$$\frac{d}{dy} \int_{g(y)}^{h(y)} f(x, y)\, dx = \int_{g(y)}^{h(y)} \frac{\partial f(x, y)}{\partial y}\, dx + f(h(y), y)\, \frac{dh(y)}{dy} - f(g(y), y)\, \frac{dg(y)}{dy},$$

donde $f(h(y), y)$ se obtiene al escribir la función $f(x, y)$ y sustituir x por $h(y)$ cada vez que aparece y lo mismo para $f(g(y), y)$. Como ilustración, si $f(x, y) = x^2y^3$, $g(y) = y$ y $h(y) = 2y$, entonces

$$\begin{aligned} \frac{d}{dy} \int_y^{2y} x^2y^3\, dx &= \int_y^{2y} 3x^2y^2\, dx + (2y)^2y^3(2) - y^2y^3(1) \\ &= 14y^5 \end{aligned}$$

para cualquier valor positivo de y.

APÉNDICE 4

Matrices y operaciones con matrices

Una **matriz** es un arreglo rectangular de números. Por ejemplo,

$$\mathbf{A} = \begin{bmatrix} 2 & 5 \\ 3 & 0 \\ 1 & 1 \end{bmatrix}$$

es una matriz de 3 × 2 (que se lee "3 por 2") pues es un arreglo rectangular de números con tres renglones y dos columnas. (Las matrices se denotan en este libro con **letras mayúsculas en negritas**.) Los números en el arreglo rectangular se conocen como **elementos** de la matriz. Por ejemplo,

$$\mathbf{B} = \begin{bmatrix} 1 & 2.4 & 0 & \sqrt{3} \\ -4 & 2 & -1 & 15 \end{bmatrix}$$

es una matriz de 2×4 cuyos elementos son 1, 2.4, 0, $\sqrt{3}$, -4, 2, -1 y 15. Así, en términos más generales,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \|a_{ij}\|$$

es una matriz de $m \times n$, donde $a_{11}, \ldots, a_{mn}$ representan los números que son los elementos de esta matriz; $\|a_{ji}\|$ es la notación abreviada para identificar la matriz cuyo elemento en el renglón i y la columna j es a_{ij} para toda $i = 1, 2, \ldots, m$ y $j = 1, 2, \ldots, n$.

OPERACIONES CON MATRICES

Como las matrices no poseen un valor numérico, no pueden sumarse, multiplicarse, etc., como si fueran números individuales. Sin embargo, a veces es necesario realizar ciertos manejos u operaciones sobre los arreglos de números. Por esta razón se han desarrollado reglas análogas a las operaciones aritméticas para realizar operaciones con matrices. Para describirlas, sean $\mathbf{A} = \|a_{ji}\|$ y $\mathbf{B} = \|b_{ji}\|$ dos matrices que tienen el mismo número de renglones y el mismo número de columnas. (Se cambiará esta restricción sobre el tamaño de **A** y **B** más adelante, cuando se analice la multiplicación de matrices.)

Se dice que **A** y **B** son *iguales*, esto es, ($\mathbf{A} = \mathbf{B}$), si y sólo si *todos* los elementos correspondientes son iguales (es decir, $a_{ij} = b_{ij}$ para toda i y j).

La operación de *multiplicar la matriz por un número* (denotado por k) se realiza al multiplicar cada elemento de la matriz por k, de modo que

$$k\mathbf{A} = \|ka_{ij}\|.$$

Por ejemplo,

$$3\begin{bmatrix} 1 & \frac{1}{3} & 2 \\ 5 & 0 & -3 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 6 \\ 15 & 0 & -9 \end{bmatrix}.$$

Para sumar **A** y **B**, simplemente se suman los elementos correspondientes, de forma que

$$\mathbf{A} + \mathbf{B} = \|a_{ij} + b_{ij}\|.$$

A manera de ilustración,

$$\begin{bmatrix} 5 & 3 \\ 1 & 6 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 3 \\ 4 & 7 \end{bmatrix}.$$

De manera similar, la *resta* se lleva a cabo como:

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-1)\mathbf{B},$$

de manera que

$$\mathbf{A} - \mathbf{B} = \|a_{ij} - b_{ij}\|.$$

Por ejemplo,

$$\begin{bmatrix} 5 & 3 \\ 1 & 6 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ -2 & 5 \end{bmatrix}.$$

Observe que, con excepción de la multiplicación por un número, todas las operaciones anteriores están definidas sólo cuando las dos matrices que intervienen son del mismo tamaño. No obstante, todas estas operaciones son directas, ya que sólo efectúan la misma comparación u operación aritmética con los elementos correspondientes de las matrices.

Existe una operación elemental adicional que no se ha definido —la **multiplicación de matrices**—, que es más complicada.

Para encontrar el elemento del renglón i y la columna j de la matriz que resulta al multiplicar la matriz **A** por la matriz **B**, es necesario multiplicar cada elemento del renglón i de **A** por el elemento correspondiente de la columna j de **B** y sumar estos productos. Para hacer esta multiplicación elemento por elemento se necesitan las siguientes restricciones sobre los tamaños de **A** y **B**:

La multiplicación de matrices **AB** está definida si y sólo si el *número de columnas* de A es igual al *número de renglones* de **B**.

Así, si **A** es una matriz de $m \times n$ y **B** es una matriz de $n \times s$, entonces su producto es

$$\mathbf{AB} = \left\| \sum_{k=1}^{n} a_{ik} b_{kj} \right\|,$$

donde este producto es una matriz de $m \times s$. Sin embargo, si **A** es una matriz de $m \times n$ y **B** es de $r \times s$, donde $n \neq r$, entonces **AB** no está definida.

Para ilustrar la multiplicación de matrices,

$$\begin{bmatrix} 1 & 2 \\ 4 & 0 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 2 & 5 \end{bmatrix} = \begin{bmatrix} 1(3) + 2(2) & 1(1) + 2(5) \\ 4(3) + 0(2) & 4(1) + 0(5) \\ 2(3) + 3(2) & 2(1) + 3(5) \end{bmatrix}$$
$$= \begin{bmatrix} 7 & 11 \\ 12 & 4 \\ 12 & 17 \end{bmatrix}.$$

Por otro lado, si se intenta multiplicar estas matrices en el orden contrario, el producto que resulta

$$\begin{bmatrix} 3 & 1 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 4 & 0 \\ 2 & 3 \end{bmatrix}$$

no está definido.

Aun cuando **AB** y **BA** estén definidos,

$$\mathbf{AB} \neq \mathbf{BA}$$

en general. Entonces, la *multiplicación de matrices* debe verse como una operación formulada especialmente, cuyas propiedades son muy diferentes de la *multiplicación aritmética*. Para comprender por qué se adoptó esta definición especial, considere el siguiente sistema de ecuaciones:

$$\begin{aligned} 2x_1 - x_2 + 5x_3 + x_4 &= 20 \\ x_1 + 5x_2 + 4x_3 + 5x_4 &= 30 \\ 3x_1 + x_2 - 6x_3 + 2x_4 &= 20. \end{aligned}$$

En lugar de escribir estas ecuaciones como se muestra aquí, se pueden escribir de manera mucho más concisa en forma matricial como

$$\mathbf{Ax} = \mathbf{b},$$

donde

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 5 & 1 \\ 1 & 5 & 4 & 5 \\ 3 & 1 & -6 & 2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 20 \\ 30 \\ 20 \end{bmatrix}.$$

Para este tipo de multiplicación se ideó la multiplicación de matrices.

Observe que la *división de matrices no* está definida.

Aunque las operaciones con matrices descritas no poseen ciertas propiedades de las operaciones aritméticas, sí satisfacen las siguientes leyes

$$\begin{aligned} \mathbf{A} + \mathbf{B} &= \mathbf{B} + \mathbf{A}, \\ (\mathbf{A} + \mathbf{B}) + \mathbf{C} &= \mathbf{A} + (\mathbf{B} + \mathbf{C}), \\ \mathbf{A}(\mathbf{B} + \mathbf{C}) &= \mathbf{AB} + \mathbf{AC}, \\ \mathbf{A}(\mathbf{BC}) &= (\mathbf{AB})\mathbf{C}, \end{aligned}$$

cuando los tamaños relativos de las matrices son tales que las operaciones indicadas están definidas.

Otro tipo de operación con matrices que no tiene análogo aritmético es la **obtención de la transpuesta**. Esta operación es nada más que el intercambio de renglones y columnas de la matriz, lo que con frecuencia es útil para realizar las operaciones de multiplicación en la forma deseada. Así, para cualquier matriz $\mathbf{A} = \|a_{ji}\|$ su transpuesta $\mathbf{A}^T$ es

$$\mathbf{A}^T = \|a_{ji}\|.$$

Por ejemplo, si

$$\mathbf{A} = \begin{bmatrix} 2 & 5 \\ 1 & 3 \\ 4 & 0 \end{bmatrix},$$

entonces

$$\mathbf{A}^T = \begin{bmatrix} 2 & 1 & 4 \\ 5 & 3 & 0 \end{bmatrix}.$$

TIPOS ESPECIALES DE MATRICES

En aritmética, 0 y 1 son números que tienen un papel especial. Existen también matrices especiales con un papel parecido en la teoría de matrices. En particular, la matriz que es análoga al 1 es la **matriz identidad I**, una matriz *cuadrada* cuyos elementos son 0 (ceros) excepto aquellos que se encuentran en la diagonal principal, los cuales son 1 (unos). Así,

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

El número de renglones o columnas de **I** puede especificarse como se desee. La analogía de **I** con el **1** se deduce del hecho de que para cualquier matriz **A**,

$$\mathbf{IA} = \mathbf{A} = \mathbf{AI},$$

donde se asigna a **I** el número adecuado de renglones y columnas para que la operación de multiplicación esté definida.

De manera similar, la matriz análoga al cero es la llamada **matriz nula** o **matriz 0**, que tiene cualquier tamaño con elementos *todos ceros*. Así,

$$\mathbf{0} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Por lo tanto, para cualquier matriz **A**,

$$\mathbf{A} + \mathbf{0} = \mathbf{A}, \qquad \mathbf{A} - \mathbf{A} = \mathbf{0} \qquad \text{y}$$
$$\mathbf{0A} = \mathbf{0} = \mathbf{A0},$$

donde **0** es del tamaño adecuado en cada caso para que la operación esté definida.

En ciertas ocasiones es útil hacer una partición de una matriz en varias matrices más pequeñas, llamadas **submatrices**. Por ejemplo, una manera posible de hacer una partición de una matriz de 3 × 4 es

$$\mathbf{A} = \left[\begin{array}{c|ccc} a_{11} & a_{12} & a_{13} & a_{14} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{array}\right] = \begin{bmatrix} a_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix},$$

donde

$$\mathbf{A}_{12} = [a_{12}, \quad a_{13}, \quad a_{14}], \qquad \mathbf{A}_{21} = \begin{bmatrix} a_{21} \\ a_{31} \end{bmatrix},$$
$$\mathbf{A}_{22} = \begin{bmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \end{bmatrix}$$

son todas submatrices. En lugar de realizar las operaciones elemento por elemento en particiones de este tipo, pueden hacerse en términos de las submatrices, siempre que las particiones sean tales que las operaciones estén definidas. Por ejemplo, si **B** es una matriz partida de 4 × 1 tal que

$$\mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ \mathbf{B}_2 \end{bmatrix},$$

entonces

$$\mathbf{AB} = \left[\begin{array}{c} a_{11}b_1 + \mathbf{A}_{12}\mathbf{B}_2 \\ \hline \mathbf{A}_{21}b_1 + \mathbf{A}_{22}\mathbf{B}_2 \end{array}\right].$$

VECTORES

Un tipo especial de matrices que tiene un papel importante en la teoría de matrices es la que tiene *un solo renglón* o *una sola columna*. Con frecuencia se hace referencia a estas matrices como **vectores**. Así,

$$\mathbf{x} = [x_1, x_2, \ldots, x_n]$$

es un **vector renglón**, y

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

es un **vector columna**. (Los vectores se denotarán aquí por **minúsculas en negritas**.) También suelen llamarse *vectores-n* para indicar que tienen *n* elementos. Por ejemplo,

$$\mathbf{x} = [1, 4, -2, \tfrac{1}{3}, 7]$$

es un "vector-5".

Un **vector nulo 0** es un vector renglón o un vector columna cuyos elementos son *todos cero*, o sea,

$$\mathbf{0} = [0, 0, \ldots, 0] \qquad \text{o} \qquad \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

(Aunque se usa el mismo símbolo **0** para el *vector nulo* y para la *matriz nula*, el contexto por lo general identifica de cuál se trata.)

Una razón por la que los vectores tienen un papel importante en la teoría de matrices es que puede hacerse una partición de cualquier matriz de $m \times n$ en m vectores renglón o bien n vectores columna y analizarse las propiedades importantes de la matriz, en términos de estos vectores. Para ampliar lo anterior, considere un conjunto de n vectores $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$, del mismo tipo (es decir, todos vectores renglón o todos vectores columna).

> **Definición:** se dice que un conjunto de vectores $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ es **linealmente dependiente** si existen m números (denotados por $c_1, c_2, \ldots, c_m$), no todos cero, tales que
>
> $$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_m\mathbf{x}_m = \mathbf{0}.$$
>
> De lo contrario, se dice que el conjunto es **linealmente independiente**.

A manera de ilustración, si $m = 3$ y

$$\mathbf{x}_1 = [1, 1, 1], \quad \mathbf{x}_2 = [0, 1, 1], \quad \mathbf{x}_3 = [2, 5, 5],$$

entonces, existen tres números, a saber, $c_1 = 2$, $c_2 = 3$ y $c_3 = -1$, tales que

$$\begin{aligned} 2\mathbf{x}_1 + 3\mathbf{x}_2 - \mathbf{x}_3 &= [2, 2, 2] + [0, 3, 3] - [2, 5, 5] \\ &= [0, 0, 0], \end{aligned}$$

entonces, $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ son linealmente dependientes. Observe que probar que son linealmente dependientes requirió encontrar tres números específicos (c_1, c_2, c_3) que hacen $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 = \mathbf{0}$, lo cual no siempre es sencillo. Observe también que esta ecuación implica que

$$\mathbf{x}_3 = 2\mathbf{x}_1 + 3\mathbf{x}_2.$$

Así, $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ serían linealmente dependientes porque uno de ellos es una combinación lineal de los otros. Sin embargo, si se cambiara $\mathbf{x}_3$ por

$$\mathbf{x}_3 = [2, 5, 6]$$

entonces $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ serían linealmente independientes puesto que es imposible expresar uno de estos vectores (por ejemplo $\mathbf{x}_3$) como una combinación lineal de los otros dos.

Definición: el **rango** de un *conjunto* de vectores es el mayor número de *vectores linealmente independientes* que pueden seleccionarse de ese conjunto.

En el ejemplo anterior, el rango del conjunto de vectores $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ era 2 (cualquier par de los vectores es linealmente independiente), pero se convirtió en 3 después de cambiar $\mathbf{x}_3$.

Definición: una **base** para un *conjunto* de vectores es una colección de vectores linealmente independientes, que se toman del conjunto, tales que todo vector del conjunto es una combinación lineal de los vectores de la colección (es decir, todo vector del conjunto es igual a la suma de ciertos múltiplos de los vectores de la colección).

Para ilustrarlo, cualquier par de vectores (como $\mathbf{x}_1$ y $\mathbf{x}_2$ anteriores) constituían una base para $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ antes de cambiar $\mathbf{x}_3$. Después de cambiar $\mathbf{x}_3$, la base se convierte en los tres vectores.

El siguiente teorema relaciona estas dos definiciones.

Teorema A4.1: Una colección de r vectores linealmente independientes, elegidos de un conjunto de vectores, es una base del conjunto si y sólo si el conjunto tiene rango r.

ALGUNAS PROPIEDADES DE MATRICES

Dados estos resultados de vectores, es posible presentar ciertos conceptos importantes de matrices.

Definición: El **rango por renglón** de una matriz es el rango de su conjunto de vectores renglón. El **rango por columna** de una matriz es el rango de su conjunto de vectores columna.

Por ejemplo, si la matriz $\mathbf{A}$ es

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 2 & 5 & 5 \end{bmatrix},$$

entonces los vectores linealmente dependientes del ejemplo anterior muestran que el rango por renglón de $\mathbf{A}$ es 2. El rango por columna de $\mathbf{A}$ también es 2. (Los primeros dos vectores columna son linealmente independientes pero el segundo vector columna menos el tercero es igual a $\mathbf{0}$.) Esto no es coincidencia, como lo indica el siguiente teorema.

Teorema A4.2: El rango por renglón y el rango por columna de una matriz son iguales.

Entonces, sólo es necesario hablar del rango de una matriz.

El concepto final que debe analizarse es el de la **inversa de una matriz**. Para cualquier número k distinto de cero, existe un recíproco o inverso, $k^{-1} = 1/k$, tal que

$$kk^{-1} = 1 = k^{-1}k.$$

¿Existe un concepto análogo que sea válido para la teoría de matrices? En otras palabras, para una matriz $\mathbf{A}$ dada, diferente de la matriz nula, ¿existe una matriz $\mathbf{A}^{-1}$ tal que

$$\mathbf{AA}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A}?$$

Si $\mathbf{A}$ no es una matriz cuadrada (esto es, si el número de renglones y de columnas son diferentes), la respuesta es *nunca*, porque estos productos de matrices necesariamente tendrían un número diferente de renglones para definir la multiplicación (de forma que no estaría definida la igualdad). No obstante, si $\mathbf{A}$ es cuadrada, entonces la respuesta es, en *ciertas circunstancias*, como se indica en la siguiente definición y en el teorema A4.3.

Definición: una matriz se llama **no singular** si su rango es igual tanto al número de renglones como al número de columnas. De lo contrario, se dice que es **singular**.

Así, sólo las matrices cuadradas pueden ser *no singulares*. Una manera útil de determinar si una matriz es no singular la proporciona el hecho de que una matriz cuadrada es no singular si y sólo si *su determinante es distinto de cero*.

Teorema A4.3:

a) Si $\mathbf{A}$ es no singular, existe una matriz no singular única $\mathbf{A}^{-1}$, llamada **inversa** de $\mathbf{A}$, tal que $\mathbf{AA}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A}$.
b) Si $\mathbf{A}$ es no singular y $\mathbf{B}$ es una matriz para la que $\mathbf{AB} = \mathbf{I}$, o $\mathbf{BA} = \mathbf{I}$, entonces $\mathbf{B} = \mathbf{A}^{-1}$.
c) Sólo las matrices no singulares tienen inversas.

A manera de ilustración de las matrices inversas, considere la matriz

$$\mathbf{A} = \begin{bmatrix} 5 & -4 \\ 1 & -1 \end{bmatrix}.$$

Observe que $\mathbf{A}$ es no singular ya que su determinante, $5(-1) -1(-4) = -1$, es distinto de cero. Por lo tanto, $\mathbf{A}$ debe tener una inversa con los elementos desconocidos

$$\mathbf{A}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Para obtener $\mathbf{A}^{-1}$ se debe usar la propiedad que

$$\mathbf{AA}^{-1} = \begin{bmatrix} 5a-4c & 5b-4d \\ a-c & b-d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

así

$$5a-4c = 1 \qquad 5b-4d = 0$$
$$a - c = 0 \qquad b - d = 1$$

Al resolver estos dos pares de ecuaciones simultanéas se encuentra que $a = 1$, $c = 1$ y $b = -4$, $d = -5$ con lo que

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & -4 \\ 1 & -5 \end{bmatrix}.$$

Por lo tanto,

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} 5 & -4 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -4 \\ 1 & -5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

y

$$\mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} 1 & -4 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 5 & -4 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

APÉNDICE 5

Tabla de una distribución normal

TABLA A5.1 **Áreas bajo la curva normal desde K_α hasta ∞**

$$\text{P\{normal estándar} > K_\alpha\} = \int_{K_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx = \alpha$$

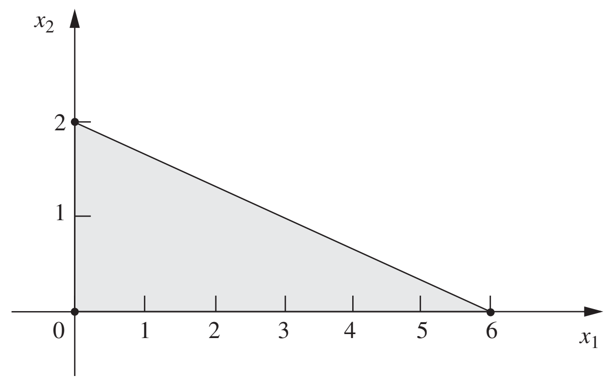
K_α	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.00990	.00964	.00939	.00914	.00889	.00866	.00842
2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
3	.00135	$.0^{3}968$	$.0^{3}687$	$.0^{3}483$	$.0^{3}337$	$.0^{3}233$	$.0^{3}159$	$.0^{3}108$	$.0^{4}723$	$.0^{4}481$
4	$.0^{4}317$	$.0^{4}207$	$.0^{4}133$	$.0^{5}854$	$.0^{5}541$	$.0^{5}340$	$.0^{5}211$	$.0^{5}130$	$.0^{6}793$	$.0^{6}479$
5	$.0^{6}287$	$.0^{6}170$	$.0^{7}996$	$.0^{7}579$	$.0^{7}333$	$.0^{7}190$	$.0^{7}107$	$.0^{8}599$	$.0^{8}332$	$.0^{8}182$
6	$.0^{9}987$	$.0^{9}530$	$.0^{9}282$	$.0^{9}149$	$.0^{10}777$	$.0^{10}402$	$.0^{10}206$	$.0^{10}104$	$.0^{11}523$	$.0^{11}260$

Fuente: F. E. Croxton, *Tables of Areas in Two Tails and in One Tail of the Normal Curve*. Derechos reservados 1949 por Prentice-Hall, Inc., Englewood Cliffs, NJ.

RESPUESTAS PARCIALES A PROBLEMAS SELECCIONADOS

CAPÍTULO 3

3.1-1. ***a*)**



3.1-5. $(x_1, x_2) = (13, 5)$; $Z = 31$.

3.1-11. ***b*)** $(x_1, x_2, x_3) = (26.19, 54.76, 20)$; $Z = 2\,904.76$.

3.2-3. ***b*)** Maximizar $\quad Z = 4\,500x_1 + 4\,500x_2$,

sujeta a

$$\begin{aligned} x_1 \qquad\qquad &\leq \quad 1 \\ x_2 &\leq \quad 1 \\ 5\,000x_1 + 4\,000x_2 &\leq 6\,000 \\ 400x_1 + \quad 500x_2 &\leq \quad 600 \end{aligned}$$

y

$$x_1 \geq 0, \qquad x_2 \geq 0.$$

3.4-3. ***a*)** *Proporcionalidad*: se cumple, puesto que se implica que una fracción fija de la dosis de radiación en un punto dado de entrada es absorbida por un área dada.

Aditividad: se cumple, puesto que se establece que la absorción de la radiación que emite rayos múltiples es aditiva.

Divisibilidad: se cumple, pues la fuerza del rayo puede tener cualquier nivel fraccionario.

Certidumbre: debido a la complejidad del análisis que se requiere para estimar los datos sobre la absorción de la radiación en los diferentes tipos de tejido, existe incertidumbre considerable sobre los datos, por lo que debe usarse el análisis de sensibilidad.

3.4-12. ***b*)** De la fábrica 1, enviar 200 unidades al cliente 2 y 200 unidades al cliente 3. De la fábrica 2, enviar 300 unidades al cliente 1 y 200 unidades al cliente 3.

3.4-13. ***c*)** $Z = \$152\,880$; $A_1 = 60\,000$; $A_3 = 84\,000$; $D_5 = 117\,600$. Todas las otras variables de decisión son 0.

3.4-15. ***b*)** Cada solución óptima tiene $Z = \$13\,330$.

3.5-2. (c, e)

Recursos	Uso de recursos por unidad de cada actividad: Actividad 1	Actividad 2	Totales		Recursos disponibles
1	2	1	10	≤	10
2	3	3	20	≤	20
3	2	4	20	≤	20
Gan. por uni.	20	30	$166.67		
Solución	3.333	3.333			

3.5-5. ***a*)** Minimizar $Z = 84C + 72T + 60A,$

sujeta a

$$90C + 20T + 40A \geq 200$$
$$30C + 80T + 60A \geq 180$$
$$10C + 20T + 60A \geq 150$$

y

$$C \geq 0, \quad T \geq 0, \quad A \geq 0.$$

CAPÍTULO 4

4.1-4. ***a*)** Las soluciones en un vértice que son *factibles* son (0, 0), (0, 1), $(\frac{1}{4}, 1)$, $(\frac{2}{3}, \frac{2}{3})$, $(1, \frac{1}{4})$ y (1, 0).

4.3-4. $(x_1, x_2, x_3) = (0, 10, 6\frac{2}{3})$; $Z = 70$.

4.6-1. ***a*), *c*)** $(x_1, x_2) = (2, 1)$; $Z = 7$.

4.6-3. ***a*), *c*), *e*)** $(x_1, x_2, x_3) = (\frac{4}{5}, \frac{9}{5}, 0)$; $Z = 7$.

4.6-9. ***a*), *b*), *d*)** $(x_1, x_2, x_3) = (0, 15, 15)$; $Z = 90$.
***c*)** Tanto en el método de la gran M como en el de dos fases, la única tabla símplex final representa una solución factible del problema real.

4.6-13. ***a*), *c*)** $(x_1, x_2) = (-\frac{8}{7}, \frac{18}{7})$; $Z = \frac{80}{7}$.

4.7-5. ***a*)** $(x_1, x_2, x_3) = (0, 1, 3)$; $Z = 7$.
***b*)** $y_1^* = \frac{1}{2}, y_2^* = \frac{5}{2}, y_3^* = 0.$ Éstos son los valores marginales de los respectivos recursos 1, 2 y 3.

CAPÍTULO 5

5.1-1. ***a*)** $(x_1, x_2) = (2, 2)$ es óptima. Otras soluciones FEV son (0, 0), (3, 0) y (0, 3).

5.1-12. $(x_1, x_2, x_3) = (0, 15, 15)$ es óptima.

5.2-2. $(x_1, x_2, x_3, x_4, x_5) = (0, 5, 0, \frac{5}{2}, 0)$; $Z = 50$.

5.3-1. ***a*)** El lado derecho es $Z = 8, x_2 = 14, x_6 = 5, x_3 = 11$.
***b*)** $x_1 = 0, 2x_1 - 2x_2 + 3x_3 = 5, x_1 + x_2 - x_3 = 3$.

CAPÍTULO 6

6.1-1. ***a*)** Minimizar $W = 15y_1 + 12y_2 + 45y_3,$

sujeta a

$$-y_1 + y_2 + 5y_3 \geq 10$$
$$2y_1 + y_2 + 3y_3 \geq 20$$

y

$$y_1 \geq 0, \quad y_2 \geq 0, \quad y_3 \geq 0.$$

6.3-1. ***c*)**

Soluciones básicas complementarias				
Problema primal			**Problema dual**	
Solución básica	**¿Factible?**	$Z = W$	**¿Factible?**	**Solución básica**
(0, 0, 20, 10)	Sí	0	No	(0, 0, −6, −8)
(4, 0, 0, 6)	Sí	24	No	$\left(1\frac{1}{5}, 0, 0, -5\frac{3}{5}\right)$
(0, 5, 10, 0)	Sí	40	No	(0, 4, −2, 0)
$\left(2\frac{1}{2}, 3\frac{3}{4}, 0, 0\right)$	Sí y óptima	45	Sí y óptima	$\left(\frac{1}{2}, 3\frac{1}{2}, 0, 0\right)$
(10, 0, −30, 0)	No	60	Sí	(0, 6, 0, 4)
(0, 10, 0, −10)	No	80	Sí	(4, 0, 14, 0)

6.3-7. ***c*)** Las variables básicas son x_1 y x_2. Las otras variables son no básicas.
***e*)** $x_1 + 3x_2 + 2x_3 + 3x_4 + x_5 = 6$, $4x_1 + 6x_2 + 5x_3 + 7x_4 + x_5 = 15$, $x_3 = 0$, $x_4 = 0$, $x_5 = 0$. La solución FEV óptima es $(x_1, x_2, x_3, x_4, x_5) = (\frac{3}{2}, \frac{3}{2}, 0, 0, 0)$.

6.4-3. Maximizar $\quad W = 8y_1 + 6y_2,$

sujeta a

$$\begin{aligned} y_1 + 3y_2 &\le 2 \\ 4y_1 + 2y_2 &\le 3 \\ 2y_1 &\le 1 \end{aligned}$$

y

$$y_1 \ge 0, \qquad y_2 \ge 0.$$

6.4-8. ***a*)** Minimizar $\quad W = 120y_1 + 80y_2 + 100y_3,$

sujeta a

$$\begin{aligned} y_2 - 3y_3 &= -1 \\ 3y_1 - y_2 + y_3 &= 2 \\ y_1 - 4y_2 + 2y_3 &= 1 \end{aligned}$$

y

$$y_1 \ge 0, \qquad y_2 \ge 0, \qquad y_3 \ge 0.$$

6.6-1. ***d*)** No es óptima puesto que $2y_1 + 3y_2 \ge 3$ se viola para $y_1^* = \frac{1}{5}$, $y_2^* = \frac{3}{5}$.
***f*)** No es óptima puesto que $3y_1 + 2y_2 \ge 2$ se viola para $y_1^* = \frac{1}{5}$, $y_2^* = \frac{3}{5}$.

6.7-1.

Inciso	Nueva solución básica $(x_1, x_2, x_3, x_4, x_5)$	¿Factible?	¿Óptima?
a)	(0, 30, 0, 0, −30)	No	No
b)	(0, 20, 0, 0, −10)	No	No
c)	(0, 10, 0, 0, 60)	Sí	Sí
d)	(0, 20, 0, 0, 10)	Sí	Sí
e)	(0, 20, 0, 0, 10)	Sí	Sí
f)	(0, 10, 0, 0, 40)	Sí	No
g)	(0, 20, 0, 0, 10)	Sí	Sí
h)	(0, 20, 0, 0, 10, $x_6 = -10$)	No	No
i)	(0, 20, 0, 0, 0)	Sí	Sí

6.7-3. $-10 \le \theta \le \frac{10}{9}$

6.7-12. ***a*)** $b_1 \ge 2$, $6 \le b_2 \le 18$, $12 \le b_3 \le 24$
***b*)** $0 \le c_1 \le \frac{15}{2}$, $c_2 \ge 2$

6.8-4. ***f*)** El intervalo permisible para permanecer óptimo de la ganancia unitaria por producir juguetes es de \$2.50 a \$5.00. El intervalo correspondiente para producir subensambles es de −\$3.00 a −\$1.50.

6.8-6. ***f)*** En el caso del inciso *a*), el cambio está dentro del incremento permisible de \$10, por lo que la solución óptima no cambia. En el del inciso *b*), el cambio está fuera de la disminución permisible de \$5, por lo que la solución óptima podría cambiar. Con respecto al inciso *c*), la suma de los porcentajes de los cambios permisibles es 250%, por lo que la regla de 100% para cambios simultáneos en los coeficientes de la función objetivo indica que la solución óptima podría cambiar.

CAPÍTULO 7

7.1-2. $(x_1, x_2, x_3) = (\frac{2}{3}, 2, 0)$ con $Z = \frac{22}{3}$ es óptima.

7.1-6. ***a)*** La nueva solución óptima es $(x_1, x_2, x_3, x_4, x_5) = (0, 0, 9, 3, 0)$ con $Z = 117$.

7.2-1. ***a, b)***

Intervalo de θ	Solución óptima	$Z(\theta)$
$0 \le \theta \le 2$	$(x_1, x_2) = (0, 5)$	$120 - 10\theta$
$2 \le \theta \le 8$	$(x_1, x_2) = \left(\frac{10}{3}, \frac{10}{3}\right)$	$\frac{320 - 10\theta}{3}$
$8 \le \theta$	$(x_1, x_2) = (5, 0)$	$40 + 5\theta$

7.2-4.

	Solución óptima		
Intervalo de θ	x_1	x_2	$Z(\theta)$
$0 \le \theta \le 1$	$10 + 2\theta$	$10 + 2\theta$	$30 + 6\theta$
$1 \le \theta \le 5$	$10 + 2\theta$	$15 - 3\theta$	$35 + \theta$
$5 \le \theta \le 25$	$25 - \theta$	0	$50 - 2\theta$

7.3-3. $(x_1, x_2, x_3) = (1, 3, 1)$ con $Z = 8$ es óptima.

CAPÍTULO 8

8.1-3. ***b)***

		Destino			
		Hoy	Mañana	Ficticio	Recursos
Origen	Dick	3.0	2.7	0	5
	Harry	2.9	2.8	0	4
	Demanda	3	4	2	

8.2-2. ***a)*** Variables básicas: $x_{11} = 4, x_{12} = 0, x_{22} = 4, x_{23} = 2, x_{24} = 0, x_{34} = 5, x_{35} = 1, x_{45} = 0; Z = 53$.
b) Variables básicas: $x_{11} = 4, x_{23} = 2, x_{25} = 4, x_{31} = 0, x_{32} = 0, x_{34} = 5, x_{35} = 1, x_{42} = 4; Z = 45$.
c) Variables básicas: $x_{11} = 4, x_{23} = 2, x_{25} = 4, x_{32} = 0, x_{34} = 5, x_{35} = 1, x_{41} = 0, x_{42} = 4; Z = 45$.

8.2-7. ***a)*** $x_{11} = 3, x_{12} = 2, x_{22} = 1, x_{23} = 1, x_{33} = 1, x_{34} = 2$; tres iteraciones para lograr la optimalidad.
b), c) $x_{11} = 3, x_{12} = 0, x_{13} = 0, x_{14} = 2, x_{23} = 2, x_{32} = 3$; ya es óptimo.

8.2-10. $x_{11} = 10, x_{12} = 15, x_{22} = 0, x_{23} = 5, x_{25} = 30, x_{33} = 20, x_{34} = 10, x_{44} = 10$; costo = \$77.30. También se tienen otras soluciones óptimas empatadas.

8.2-11. ***b)*** Sea x_{ij} el envío desde la planta *i* al centro de distribución *j*. Entonces $x_{13} = 2, x_{14} = 10, x_{22} = 9, x_{23} = 8, x_{31} = 10, x_{32} = 1$; costo = \$20 200.

8.3-4. ***a)***

		Tarea				
		Dorso	Pecho	Mariposa	Libre	Ficticio
	Carl	37.7	43.4	33.3	29.2	0
	Chris	32.9	33.1	28.5	26.4	0
Asignado	David	33.8	42.2	38.9	29.6	0
	Tony	37.0	34.7	30.4	28.5	0
	Ken	35.4	41.8	33.6	31.1	0

CAPÍTULO 9

9.3-3. ***a***) $O \to A \to B \to D \to T$ o $O \to A \to B \to E \to D \to T$, con longitud = 16.

9.4-1. ***a***) $\{(O, A); (A, B); (B, C); (B, E); (E, D); (D, T)\}$, con longitud = 18.

9.5-1.

Arco	(1, 2)	(1, 3)	(1, 4)	(2, 5)	(3, 4)	(3, 5)	(3, 6)	(4, 6)	(5, 7)	(6, 7)
Flujo	4	4	1	4	1	0	3	2	4	5

9.8-3. ***a***) Ruta crítica: Inicio $\to A \to C \to E \to$ Final
Duración total = 12 semanas

b) Nuevo plan:

Actividad	Duración	Costo
A	3 semanas	$54 000
B	3 semanas	$65 000
C	3 semanas	$68 666
D	2 semanas	$41 500
E	2 semanas	$80 000

Se ahorran $7 834 con este programa acelerado.

CAPÍTULO 10

10.3-2.

	Tienda		
	1	**2**	**3**
Asignaciones	1 3	2 2	2 0

10.3-7. ***a***)

Fase	***a***)	***b***)
1	2*M*	2.945*M*
2	1*M*	1.055*M*
3	1*M*	0
Porcentaje del mercado	6 %	6.302 %

10.3-11. $x_1 = -2 + \sqrt{13} \approx 1.6056$, $x_2 = 5 - \sqrt{13} \approx 1.3944$; $Z = 98.233$.

10.4-3. Producir 2 en la primera corrida de producción; si ninguno es aceptable, producir 3 en la segunda corrida. Costo esperado = $573.

CAPÍTULO 11

11.1-2. ***a***) Minimizar $Z = 4.5x_{em} + 7.8x_{ec} + 3.6x_{ed} + 2.9x_{el} + 4.9x_{sm} + 7.2x_{sc} + 4.3x_{sd} + 3.1x_{sl}$,

sujeta a

$$\begin{aligned} x_{em} + x_{ec} + x_{ed} + x_{el} &= 2 \\ x_{sm} + x_{sc} + x_{sd} + x_{sl} &= 2 \\ x_{em} + x_{sm} &= 1 \\ x_{ec} + x_{sc} &= 1 \\ x_{ed} + x_{sd} &= 1 \\ x_{el} + x_{sl} &= 1 \end{aligned}$$

y

todas las x_{ij} son binarias.

11.3-1. ***b*)**

Restricción	Producto 1	Producto 2	Producto 3	Producto 4	Totales		Lado derecho modificado	Lado derecho original
Primera	5	3	6	4	6 000	≤	6 000	6 000
Segunda	4	6	3	5	12 000	≤	105 999	6 000
Ingreso marginal	$70	$60	$90	$80	$80 000			
Solución	0	2 000	0	0				
	≤	≤	≤	≤				
	0	9 999	0	0				
¿Costo preparación?	0	1	0	0	1	≤	2	
Costo inicio	$50 000	$40 000	$70 000	$60 000				

Restricciones de contingencia:

Producto 3:	0	≤	1	:Producto 1 o 2
Producto 4:	0	≤	1	:Producto 1 o 2

Qué restricción (0 = primera, 1 = segunda):	0

11.3-5. ***b*, *d*)** (grande, mediano, chico) = (14, 0, 16), con una ganancia de $95.6 millones.

11.4-3. ***b*)**

Restricción	Producto 1	Producto 2	Producto 3	Totales		Lado derecho
Fresadora	9	3	5	498	≤	500
Torno	5	4	0	349	≤	350
Rectificadora	3	0	2	135	≤	150
Pot. de ventas	0	0	1	0	≤	20
Gan. unitaria	50	20	25	$2 870		
Solución	45	31	0			
	≤	≤	≤			
	999	999	0			
¿Producir?	1	1	0	2	≤	2

11.4-5. ***a*)** Sea $x_{ij} = \begin{cases} 1 & \text{si el arco } i \rightarrow j \text{ es incluido en la trayectoria más corta} \\ 0 & \text{de otra manera} \end{cases}$

Alternativas mutuamente excluyentes: en cada columna de arcos se incluye exactamente un arco en la ruta más corta. Decisiones contingentes: la ruta más corta sale del nodo *i* sólo si entra al nodo *i*.

11.5-2. ***a*)** $(x_1, x_2) = (2, 3)$ es óptima.
***b*)** Ninguna de las soluciones factibles redondeadas es óptima para el problema de programación entera.

11.6-1. $(x_1, x_2, x_3, x_4, x_5) = (0, 0, 1, 1, 1)$, con $Z = 6$.

11.6-7. ***b*)**

Tarea	1	2	3	4	5
Asignado	1	3	2	4	5

11.6-9. $(x_1, x_2, x_3, x_4) = (0, 1, 1, 0)$, con $Z = 36$.

11.7-2. ***a*), *b*)** $(x_1, x_2) = (2, 1)$ es óptima.

11.8-1. ***a*)** $x_1 = 0, x_3 = 0$

CAPÍTULO 12

12.2-7. ***a*)** Cóncava.

12.4-1. ***a*)** Solución aproximada = 1.0125.

12.5-3. La solución exacta es $(x_1, x_2) = (2, -2)$.

12.5-7. ***a*)** La solución aproximada es $(x_1, x_2) = (0.75, 1.5)$.

12.6-3.

$$\begin{aligned}
-4x_1^3 - 4x_1 - 2x_2 + 2u_1 + u_2 &= 0 \quad (\text{o} \leq 0 \text{ si } x_1 = 0).\\
-2x_1 - 8x_2 + u_1 + 2u_2 &= 0 \quad (\text{o} \leq 0 \text{ si } x_2 = 0).\\
-2x_1 - x_2 + 10 &= 0 \quad (\text{o} \leq 0 \text{ si } u_1 = 0).\\
-x_1 - 2x_2 + 10 &= 0 \quad (\text{o} \leq 0 \text{ si } u_2 = 0).
\end{aligned}$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad u_1 \geq 0, \quad u_2 \geq 0.$$

12.6-6. $(x_1, x_2) = (1, 2)$ no puede ser óptima.

12.6-8. ***a*)** $(x_1, x_2) = (1 - 3^{-1/2}, 3^{-1/2})$.

12.7-2. ***a*)** $(x_1, x_2) = (2, 0)$ es óptima.

***b*)** Minimizar $\quad Z = z_1 + z_2$,

sujeta a

$$\begin{aligned}
2x_1 + u_1 - y_1 + z_1 &= 8\\
2x_2 + u_1 - y_2 + z_2 &= 4\\
x_1 + x_2 + v_1 &= 2
\end{aligned}$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad u_1 \geq 0, \quad y_1 \geq 0, \quad y_2 \geq 0, \quad v_1 \geq 0, \quad z_1 \geq 0,$$
$$z_2 \geq 0.$$

12.8-2. ***b*)** Maximizar $\quad Z = 3x_{11} - 3x_{12} - 15x_{13} + 4x_{21} - 4x_{23}$,

sujeta a

$$\begin{aligned}
x_{11} + x_{12} + x_{13} + 3x_{21} + 3x_{22} + 3x_{23} &\leq 8\\
5x_{11} + 5x_{12} + 5x_{13} + 2x_{21} + 2x_{22} + 2x_{23} &\leq 14
\end{aligned}$$

y

$$0 \leq x_{ij} \leq 1, \quad \text{para } i = 1, 2, 3; j = 1, 2, 3.$$

12.9-8. ***a*)** $(x_1, x_2) = \left(\frac{1}{3}, \frac{2}{3}\right)$.

12.9-14. ***a*)** $P(x; r) = -2x_1 - (x_2 - 3)^2 - r\left(\frac{1}{x_1 - 3} + \frac{1}{x_2 - 3}\right)$.

***b*)** $(x_1, x_2) = \left[3 + \left(\frac{r}{2}\right)^{1/2}, 3 + \left(\frac{r}{2}\right)^{1/3}\right]$

CAPÍTULO 13

13.2-2. La mejor solución que se encontró tiene ligaduras AC, BC, CD y DE.

13.4-2. ***a*)** En el caso del primer hijo, las opciones de la primera ligadura son 1-2, 1-8, 1-5 y 1-4, por lo que los números aleatorios 0.09656 y 0.96657 indican que se debe elegir la ligadura 1-2 y que no ocurren mutaciones. Las opciones de la segunda ligadura son, entonces, 2-3, 2-8 y 2-4, y así de manera sucesiva. Ocurre una mutación en la quinta ligadura. El primer hijo completo es 1-2-8-5-6-4-7-3-1.

CAPÍTULO 14

14.2-2. ***a*)** Jugador 1: estrategia 2; jugador 2: estrategia 1.

14.2-7. ***a*)** Político 1: tema 2; político 2: tema 2.

***b*)** Político 1: tema 1; político 2: tema 2.

14.4-4. $(x_1, x_2) = (\frac{2}{5}, \frac{3}{5})$; $(y_1, y_2, y_3) = (\frac{1}{5}, 0, \frac{4}{5})$; $v = \frac{8}{5}$.

14.5-3. ***a*)** Maximizar $\quad x_4$,

sujeta a

$$\begin{aligned}
5x_1 + 2x_2 + 3x_3 - x_4 &\geq 0\\
4x_2 + 2x_3 - x_4 &\geq 0\\
3x_1 + 3x_2 - x_4 &\geq 0\\
x_1 + 2x_2 + 4x_3 - x_4 &\geq 0\\
x_1 + x_2 + x_3 &= 1
\end{aligned}$$

y

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0.$$

CAPÍTULO 15

15.2-2. ***a***)

	Estado de la naturaleza	
Alternativa	**Vender 10 000**	**Vender 100 000**
Construir computadoras	0	54
Vender los derechos	15	15

c) Sea p = probabilidad *a priori* de vender 10 000. Deben construir cuando $p \leq 0.722$ y vender cuando $p > 0.722$.

15.2-4. ***c***) Warren debe hacer una inversión contracíclica.

15.2-8. Ordenar 25.

15.3-2. ***a***) VEIP = PE (con información perfecta) – PE (sin más información) = 34.5 – 27 = $7.5 millones.

d)

Datos:		P (Resultado | Estado) Resultado	
Estado de la naturaleza	**Probabilidad *a priori***	**Vender 10 000**	**Vender 100 000**
Vender 10 000	0.5	0.666666667	0.333333333
Vender 100 000	0.5	0.333333333	0.666666667

Probabilidad *a posteriori*		P (Estado | Resultado) Estado de la naturaleza	
Resultado	***P* (Resultado)**	**Vender 10 000**	**Vender 100 000**
Vender 10 000	0.5	0.666666667	0.333333333
Vender 100 000	0.5	0.333333333	0.666666667

15.3-4. ***b***) VEIP = PE (con información perfecta) – PE (sin más información) = 53 – 35 = $18

c) Betsy puede gastar hasta $18 para obtener más información.

15.3-8. ***a***) Hasta $230 000

b) Ordenar 25.

15.3-9. ***a***)

	Estado de la naturaleza		
Alternativa	**Riesgo malo**	**Riesgo promedio**	**Riesgo bueno**
Extender el crédito	−15 000	10 000	20 000
No extender el crédito	0	0	0
Probabilidades *a priori*	0.2	0.5	0.3

c) VEIP = PE (con información perfecta) – PE (sin más información) = 11 000 – 8 000 = $3 000. Este resultado indica que no se debe utilizar la organización de clasificación de crédito.

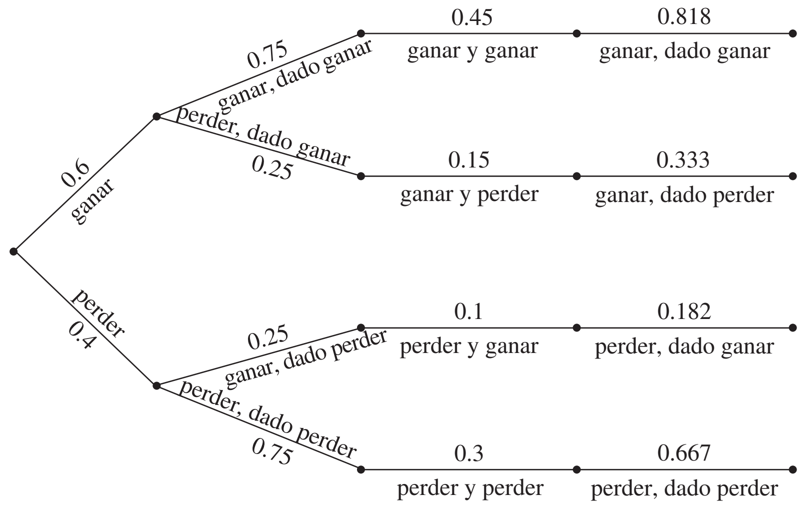
15.3-13. ***a***) Suponga moneda 1.

b) Cara: moneda 2; cruz: moneda 1.

15.4-2. La política óptima es no hacer investigación de mercado y construir computadoras.

15.4-4. ***c***) VEIP = PE (con información perfecta) – PE (sin más información) = 1.8 – 1 = $800 000

d) Probabilidades *a priori* P (estado) | Probabilidades condicionales P (resultado|estado) | Probabilidades conjuntas P (estado y resultado) | Probabilidades *a posteriori* P (estado|resultado)



f) Leland University debe contratar a William. Si él predice una temporada ganadora deben realizar la campaña; si predice una temporada perdedora, deben suspenderla.

15.5-7. ***a***) Seleccione introducir el producto nuevo (la ganancia esperada es de $12.5 millones).
b) $7.5 millones.
c) La política óptima es no hacer la prueba e introducir el nuevo producto.
f) Ambas gráficas indican que el pago esperado es sensible a los dos parámetros, pero es más sensible a cambios en la ganancia si tiene éxito que a los cambios en la pérdida si no lo tiene.

15.6-2. ***a***) Escoger no comprar el seguro (la ganancia esperada es de $249 840).
b) U(asegurado) = 499.82
U(sin seguro) = 499.8
La política óptima es comprar el seguro.

15.6-4. $U(10) = 9$

CAPÍTULO 16

16.3-3. ***c***) $\pi_0 = \pi_1 = \pi_2 = \pi_3 = \pi_4 = \frac{1}{5}$.

16.4-1. ***a***) Todos los estados pertenecen a la misma clase recurrente.

16.5-7. ***a***) $\pi_0 = 0.182$, $\pi_1 = 0.285$, $\pi_2 = 0.368$, $\pi_3 = 0.165$.
b) 6.50

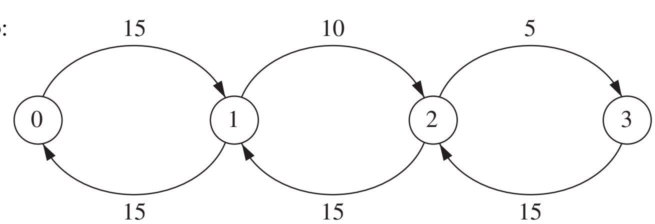
CAPÍTULO 17

17.2-1. Fuente de entrada: población con pelo; clientes: clientes que necesitan corte de pelo; y así sucesivamente para la cola, la disciplina de la cola y el mecanismo de servicio.

17.2-2. ***b***) $L_q = 0.375$
d) $W - W_q = 24.375$ minutos.

17.4-2. ***c***) 0.0527

17.5-5. ***a***) Estado:



c) $P_0 = \frac{9}{26}, P_1 = \frac{9}{26}, P_2 = \frac{3}{13}, P_3 = \frac{1}{13}$.
d) $W = 0.11$ horas.

17.5-8. ***b***) $P_0 = \frac{2}{5}, P_n = (\frac{3}{5})(\frac{1}{2})^n$
c) $L = \frac{6}{5}, L_q = \frac{3}{5}, W = \frac{1}{25}, W_q = \frac{1}{50}$

17.6-2. ***a***) $P_0 + P_1 + P_2 + P_3 + P_4 = 0.96875$ o 97% del tiempo.

17.6-21. ***a***) Tiempo de espera esperado combinado = 0.211.
c) Un tiempo esperado de proceso de 3.43 minutos ocasionará que los tiempos de espera esperados de los dos procedimientos sean los mismos.

17.6-26. ***a***) 0.429

17.6-32. ***a***) tres máquinas
b) tres operadores

17.7-1. ***a***) W_q (exponencial) $= 2W_q$ (constante) $= \frac{8}{5}W_q$ (Erlang).
b) W_q (nuevo) $= \frac{1}{2}W_q$ (antiguo) y L_q (nuevo) $= L_q$ (antiguo) para todas las distribuciones.

17.7-6. ***a***, ***b***) Bajo la política actual un avión pierde 1 día de vuelo en lugar de los 3.25 días bajo la política propuesta.
Con la política actual un avión pierde tiempo de vuelo por día en lugar de los 0.8125 de avión.

17.7-9.

Distribución de servicio	P_0	P_1	P_2	L
Erlang	0.561	0.316	0.123	0.561
Exponencial	0.571	0.286	0.143	0.571

17.8-1. ***a***) Este sistema es un ejemplo de un sistema de colas con prioridades sin interrupción.
c) $\dfrac{W_q \text{ para pasajeros de primera clase}}{W_q \text{ para pasajeros de clase turista}} = \dfrac{0.033}{0.083} = 0.4$

17.8-4. ***a***) $W = \frac{1}{2}$
b) $W_1 = 0.20, W_2 = 0.35, W_3 = 1.10$
c) $W_1 = 0.125, W_2 = 0.3125, W_3 = 1.250$

17.10-2. Cuatro cajas registradoras.

CAPÍTULO 18

18.3-1. ***a***) $t = 1.83, Q = 54.77$
b) $t = 1.91, Q = 57.45, S = 52.22$

18.3-3. ***a***)

Datos

d =	676	(demanda/año)
K =	$75	(costo fijo)
h =	$600.00	(costo uni. de mant.)
L =	3.5	(tiempo de entrega en días)
WD =	365	(días hábiles/año)

Resultados

Punto de reorden =	6.5
Costo fijo anual =	$10 140
Costo de mantener anual =	$ 1 500
Costo variable total =	$11 640

Decisión

Q =	5	(cantidad a ordenar)

d) **Datos**

d =	676	(demanda/año)
K =	$75	(costo fijo)
h =	$600	(costo uni. de mant.)
L =	3.5	(tiempo de entrega en días)
WD =	365	(días hábiles/año)

Resultados

Punto de reorden =	6.48
Costo fijo anual =	$3 900
Costo de mantener anual =	$3 900
Costo variable total =	$7 800

Decisión

Q =	13	(cantidad a ordenar)

Los resultados son los mismos que los que se obtuvieron en el inciso *c*).

f) Número de órdenes por año = 52.
PDR = 6.5 – nivel de inventario cuando se coloca la orden.

g) La política óptima reduce el costo variable total de inventario en $3 840 por año, lo cual es una reducción de 33%.

18.3-6. ***a)*** $h = \$3$ por mes que es 15% del costo de adquisición.
c) El punto de reorden es 10.
d) *PDR* = 5 martillos, lo cual suma $20 a su CVT (5 martillos × $4 del costo de mantener).

18.3-7. $t = 3.26$, $Q = 26\ 046$, $S = 24\ 572$

18.3-12. ***a)*** óptima $Q = 500$.

18.4-4. Producir 3 unidades en el periodo 1 y 4 unidades en el periodo 3.

18.6-6. ***b)*** Ground Chuck: $R = 145$.
Chuck Wagon: $R = 829$.
c) Ground Chuck: inventario de seguridad = 45.
Chuck Wagon: inventario de seguridad = 329.
f) Ground Chuck: $39 378.71.
Chuck Wagon: $41 958.61.
Jed debe elegir a Ground Chuck como su proveedor.
g) Si Jed quiere utilizar la carne dentro del mes de recibirla, la mejor opción es Ground Chuck. La cantidad a ordenar a Ground Chuck es de alrededor del suministro de un mes, mientras que la cantidad a ordenar óptima a Chuck Wagon equivale a una provisión aproximada de tres meses.

18.7-5. ***a)*** Nivel de servicio óptimo = 0.667
c) $Q^* = 500$
d) La probabilidad de faltantes es 0.333
e) Nivel de servicio óptimo = 0.833

CAPÍTULO 19

19.2-2. ***c)*** Usar servicio lento cuando no hay ninguno o un solo cliente y el servicio rápido cuando dos clientes están presentes.

19.2-3. ***a)*** Los estados posibles del automóvil son abollado y no abollado.
c) Cuando el automóvil no esté abollado, estacionarlo en un espacio en la calle. Cuando esté abollado, llevarlo a reparación.

19.2-5. ***c)*** Estado 0: intentar servicio as; estado 1: intentar un "globo".

19.3-2. ***a)*** Minimizar $Z = 4.5y_{02} + 5y_{03} + 50y_{14} + 9y_{15}$,
sujeta a

$$y_{01} + y_{02} + y_{03} + y_{14} + y_{15} = 1$$

$$y_{01} + y_{02} + y_{03} - \left(\frac{9}{10}y_{01} + \frac{49}{50}y_{02} + y_{03} + y_{14}\right) = 0$$

$$y_{14} + y_{15} - \left(\frac{1}{10}y_{01} + \frac{1}{50}y_{02} + y_{15}\right) = 0$$

y

todas las $y_{ik} \geq 0$.

19.3-4. ***a***) Minimizar $Z = -\frac{1}{8}y_{01} + \frac{7}{24}y_{02} + \frac{1}{2}y_{11} + \frac{5}{12}y_{12}$,

sujeta a

$$y_{01} + y_{02} - \left(\frac{3}{8}y_{01} + y_{11} + \frac{7}{8}y_{02} + y_{12}\right) = 0$$
$$y_{11} + y_{12} - \left(\frac{5}{8}y_{01} + \frac{1}{8}y_{02}\right) = 0$$
$$y_{01} + y_{02} + y_{11} + y_{12} = 1$$

y

$y_{ik} \geq 0$ para $i = 0, 1; k = 1, 2.$

19.4-2. Automóvil no abollado: estacionarlo en un espacio en la calle. Automóvil abollado: repararlo.

19.4-4. Estado 0: intentar un servicio as. Estado 1: intentar un "globo".

19.5-1. Rechazar oferta de $600; aceptar cualquiera de las otras dos.

19.5-2. ***a***) Minimizar $Z = 60(y_{01} + y_{11} + y_{21}) - 600y_{02} - 800y_{12} - 1\ 000y_{22}$,

sujeta a

$$y_{01} + y_{02} - (0.95)\left(\frac{5}{8}\right)(y_{01} + y_{11} + y_{21}) = \frac{5}{8}$$

$$y_{11} + y_{12} - (0.95)\left(\frac{1}{4}\right)(y_{01} + y_{11} + y_{21}) = \frac{1}{4}$$

$$y_{21} + y_{22} - (0.95)\left(\frac{1}{8}\right)(y_{01} + y_{11} + y_{21}) = \frac{1}{8}$$

y

$y_{ik} \geq 0$ para $i = 0, 1,2; k = 1, 2.$

19.5-3. Después de tres iteraciones la aproximación es, de hecho, la política óptima que se presenta en el problema 19.5-1.

19.5-11. En los periodos 1 a 3 no hacer nada cuando la máquina está en los estados 0 o 1; reparación general cuando la máquina está en el estado 2 y reemplazo cuando está en el estado 3. En el periodo 4, no hacer nada si la máquina está en los estados 0, 1 o 2, y reemplazo cuando está en el estado 3.

CAPÍTULO 20

20.1-1. ***b***) Haga que los números del 0.000 a 0.5999 correspondan a los strikes y los números 0.6000 a 0.9999 a las bolas. Las observaciones aleatorias de las pichadas son 0.7520 = bola, 0.4184 = strike, 0.4189 = strike, 0.9559 = bola y 0.1403 = strike.

20.1-10. ***b***) Use $\lambda = 4$ y $\mu = 5$.

i) Las respuestas pueden variar. La opción de entrenar a los dos mecánicos actuales baja significativamente el tiempo de espera de los autos alemanes, pero sin efecto de importancia en el tiempo de espera de los autos japoneses y sin tener el costo adicional de un tercer mecánico. Agregar el tercer mecánico reduce aún más la espera promedio de los autos alemanes, pero con el costo adicional de tener un tercer mecánico.

20.3-1. ***a***) 5, 8, 1, 4, 7, 0, 3, 6, 9, 2.

20.4-2. ***b***) $F(x) = 0.0965$ donde $x = -25.18$
$F(x) = 0.5692$ donde $x = 18.46$
$F(x) = 0.6658$ donde $x = 23.29$

20.4-6. ***a***) Ejemplo de una réplica.

Resumen de resultados

¿Ganar? (1 = sí, 0 = no)	0
Número de lanzamientos =	3

Lanzamientos simulados

Lanzamiento	Dado 1	Dado 2	Suma
1	4	2	6
2	3	2	5
3	6	1	7
4	5	2	7
5	4	4	8
6	1	4	5
7	2	6	8

Resultados

¿Ganar?	¿Perder?	¿Continuar?
0	0	Sí
0	0	Sí
0	1	No
NA	NA	No
NA	NA	No
NA	NA	No
NA	NA	No

ÍNDICE ONOMÁSTICO

ÍNDICE ANALÍTICO

Los números de página seguidos de una “n” indican que la entrada se encuentra en las notas

G

H

I

J

K

L

M

S

U

V